

Universidade
Federal de
Santa Catarina

Programa de
Pós-Graduação
em Ciência da
Computação

www.ppgcc.inf.ufsc.br/

Campus
Universitário
Trindade
Florianópolis -
SC

Dissertação apresentada ao programa de Pós-Graduação em
Ciência da Computação, Departamento de Informática e
Estatística, do Centro Tecnológico da Universidade Federal de
Santa Catarina, como requisito para a obtenção do título de
Mestre em Ciência da Computação.

Orientador: Prof. Mauro Roisenberg, Dr.

Florianópolis, 2010
FF

Intervalo de Predição em Redes RBF: Análise e Proposta de Extensão
Abner Cardoso Rodrigues Neto

Intervalo de Predição em Redes RBF: Análise e
Proposta de Extensão

Abner Cardoso Rodrigues Neto

Dissertação de
Mestrado

Prof. Mauro
Roisenberg, Dr.



UNIVERSIDADE FEDERAL DE SANTA CATARINA PROGRAMA DE
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ABNER CARDOSO RODRIGUES NETO

INTELIGÊNCIA COMPUTACIONAL

Intervalo de Predição em Redes RBF: Análise e Proposta de Extensão

Orientador:
Prof. Mauro Roisenberg, Dr.

Florianópolis, 2010

Catálogo na fonte pela Biblioteca Universitária
da
Universidade Federal de Santa Catarina

R696i Rodrigues Neto, Abner Cardoso
Intervalo de Predição em reded RBF [dissertação] : análise
e proposta de extensão / Abner Cardoso Rodrigues Neto ;
orientador, Mauro Roisenberg. - Florianópolis, SC, 2010.
57 p.: il., grafs., tabs.

Dissertação [Mestrado] - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Computação.

Inclui referências

1. Ciência da computação. 2. Redes neurais [Computação].
3. Predição. 4. Redes de computadores - Medidas de segurança.
I. Roisenberg, Mauro. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Ciência da Computação.
III. Título.

CDU 681

ABNER CARDOSO RODRIGUES NETO

INTERVALO DE PREDIÇÃO EM REDES RBF: ANÁLISE E PROPOSTA
DE EXTENSÃO

Esta dissertação foi julgada adequada para obtenção do Título de "Mestre em
Ciência da Computação", e aprovada em sua forma final pelo Programa de
Pós-Graduação em Ciência da Computação.

Florianópolis, 12 de Novembro de 2010

Prof. Mário Antônio Ribeiro Dantas, PhD.
Coordenador do Curso

Banca Examinadora:

Prof. Mauro Roisenberg, Dr.
Orientador

Prof. Gustavo Augusto Lima de Campos, Dr.
Universidade Estadual do Ceará

Prof. Ricardo Azambuja Silveira, Dr.
Universidade Federal de Santa Catarina

Prof. Pedro Alberto Barbeta, Dr.
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Agradeço aos meus pais e meus avós pelo o apoio e compreensão. Ao Professor Mauro pela paciência e orientação durante esses anos de mestrado e ao Professor Gustavo pela confiança e palavras de incentivo que me fizeram optar pela pesquisa científica.

RESUMO

RODRIGUES NETO, Abner Cardoso. Intervalo de Predição em Redes RBF: Análise e Proposta de Extensão. Florianópolis, 2010. 57 f. Dissertação (Mestrado) Programa de Pós-Graduação em Ciência da Computação, UFSC, Florianópolis – SC.

Redes Neurais são amplamente empregadas em problemas de classificação e regressão, porém os modelos mais comuns fornecem apenas a estimação de regressão sem nenhuma medida de confiança associada à saída da rede. Medidas de desempenho global como o Erro Médio Quadrático não são capazes de reconhecer regiões onde a resposta da rede possa estar contaminada com incertezas, devido ao ruído presente nos dados ou à baixa densidade de dados de treinamento nessas regiões.

Incorporar medidas de confiança na saída da rede, como intervalos de predição, valida a regressão e auxilia tomadores de decisão a estabelecerem critérios de risco, necessários em muitas aplicações práticas. Entretanto, existe uma série de restrições para o cálculo do Intervalo de Predição nas redes neurais, que são difíceis de serem cumpridas em problemas reais.

Neste trabalho, estudou-se as medidas de confiança fornecida pela rede de função de base radial, algumas das suas deficiências foram tratadas com o objetivo de obter medidas de confiança mais satisfatórias e com menos restrições sobre o modelo, que possam ajudar os tomadores de decisão em aplicações reais.

PALAVRAS-CHAVE: Redes Neurais, Intervalo de Predição, RBF

ABSTRACT

RODRIGUES NETO, Abner Cardoso. Prediction Interval in RBF Networks: Analysis and Extension . Florianópolis, 2010. 57 f. Thesis (Masters) Post-Graduate in Computer Science, UFSC, Florianópolis – SC.

Neural networks are widely employed in problems of classification and regression, but the most common models provide only the estimation of regression without any measure of confidence associated with the network output. Global performance measures such as Mean Square Error are not able to recognize areas where the response of the network may be contaminated with uncertainties due to noise present in the data or low density of training data in these regions.

Incorporating confidence measures in the output of the network, such as prediction intervals, validates the estimation and assists decision makers to establish risk criteria, necessary in many practical applications. However there are a number of constraints for the calculation of the prediction interval in neural networks, which are difficult to enforce in real problems.

In this work, we studied the confidence measures provided by radial basis function networks, some of its deficiencies were treated with the aim of obtaining more satisfactory measures of confidence and with fewer restrictions on the model, that may help decision makers in real applications.

KEYWORDS: Neural networks, prediction interval, RBF

LISTA DE FIGURAS

2.1	Modelo de um neurônio, as entradas são ponderadas por pesos que são somadas pelo somador e depois servem como entrada para a função de ativação.	17
2.2	Gráfico das funções de base radial mais usadas na redes RBF, todas foram calculadas com raio 1 e centros 0.	19
2.3	Diagrama da arquitetura das redes RBF.	20
2.4	Aproximação da rede RBF utilizando raios calculados segundo a heurística dos vizinhos mais próximos.	23
2.5	Aproximação da rede RBF utilizando raios bem maiores (multiplicados por 1000).	23
2.6	Aproximação da rede RBF utilizando raios bem menores (divididos por 5).	24
2.7	Arquitetura da rede VInet.	28
2.8	Histograma dos dados de treinamento com densidade uniforme.	29
2.9	Saída da rede, intervalos de confiança e saída desejada.	30
2.10	Intervalo de predição para os dados de teste.	30
2.11	Esquema ilustrando como se utiliza a densidade como medida de confiança.	31
2.12	Histograma dos dados de treinamento de dados não uniformes.	33
2.13	Densidade calculada pela rede com 310 unidades escondidas.	33
2.14	Densidade calculada pela rede com 4 unidades escondidas.	34
2.15	Saída da rede quando utiliza-se 310 unidades escondidas.	34
2.16	Saída da rede quando utiliza-se 4 unidades escondidas.	35
3.1	Elementos básicos de uma rede SOM, os l neurônios da grade estão dispostos em um formato retangular, no instante n , um padrão x é apresentado a rede e dentre todos os neurônios é escolhido o vencedor v . A vizinhança é definida por h_{vi} centrada no vencedor e apenas esses neurônios possuem suas conexões ajustadas de forma que se aproximem do padrão de entrada.	37
3.2	Posição dos pesos dos neurônios de uma SOM treinada com dados não uniformemente distribuídos. Os neurônios são representados pelos círculos e os pontos de treinamento pelos caractere x	38
4.1	IP para o Conjunto 1, dados de treinamento com 3 regiões com ruído diferente.	45

	10
4.2 IP para o o Conjunto 5, dados de treinamento com diferentes níveis de ruído e densidade, a região com mais ruído possui menor densidade.	45
4.3 IP para o Conjunto 7, ruído constante mas regiões com densidade diferente.	46
4.4 Saída desejada no problema de caracterização de reservatório. . . .	48
4.5 Conjunto de treinamento no problema de caracterização de reservatório. Os círculos mostram as regiões com diferentes níveis de ruído.	49
4.6 IP para o problema de caracterização de reservatórios.	49
4.7 IP da E-VInet exibindo como uma seção vertical.	50

LISTA DE TABELAS

- 4.1 Resumo da densidade e ruído nos dados usados no primeiro teste. A variação da densidade significa como o conjunto de treinamento está distribuído em cada um dos três intervalos do espaço de entrada. A fração é a razão dos dados que pertence a cada intervalo. 43
- 4.2 Número de dados de treinamento em cada conjunto. 43
- 4.3 PCIP para cada conjunto de treinamento. 44
- 4.4 PCIP para o segundo experimento, o primeiro teste possui 2104 pontos escolhidos aleatoriamente, o segundo teste possui 1052 pontos também escolhidos aleatoriamente e o terceiro possui 740 pontos. 47
- 4.5 PCIP para o problema de caracterização de reservatórios, o conjunto de treinamento possui 748 pontos e o de teste possui 4806. . . . 48

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONTEXTO DO TRABALHO	13
1.2	OBJETIVOS GERAL E ESPECÍFICOS	15
1.3	JUSTIFICATIVA E RELEVÂNCIA	15
1.4	CONTRIBUIÇÕES RELEVANTES	15
1.5	ORGANIZAÇÃO DA DISSERTAÇÃO	16
2	MEDIDAS DE CONFIANÇA PARA REDES DE FUNÇÃO DE BASE RADIAL	17
2.1	TREINAMENTO DAS REDES RBF	19
2.2	COMPARAÇÃO COM REDES MLP	22
2.3	REDE ÍNDICE DE VALIDADE	25
2.4	LIMITAÇÕES DO INTERVALO DE PREDIÇÃO DA VINET	28
2.4.1	Influência da Densidade no Intervalo de Predição	29
2.4.2	Cálculo da Densidade na VInet	31
3	PROPOSTA PARA EXTENSÃO DO MODELO VINET	36
3.1	CALCULANDO DENSIDADE COM SOM	36
3.2	COMBINANDO DENSIDADE COM INTERVALO DE PREDIÇÃO	40
4	TESTES E RESULTADOS	41
4.1	METODOLOGIA	41
4.2	EXPERIMENTO 1	42
4.3	EXPERIMENTO 2	46
4.4	EXPERIMENTO 3	47
5	CONCLUSÕES	51
5.1	LIMITAÇÕES E TRABALHOS FUTUROS	51
	REFERÊNCIAS	57

1 INTRODUÇÃO

1.1 CONTEXTO DO TRABALHO

Devido a sua grande capacidade representacional, as redes neurais têm sido largamente empregadas como uma ferramenta para resolver problemas de regressão e classificação (HAYKIN, 1998), (CICHOCKI; UNBEHAUEN, 1993), (J. ROGERS J.A.; R., 1991), (LORENZETTI, 1992). Os modelos mais comuns de redes neurais fornecem apenas a estimação da regressão, sem uma medida de confiança ou um Intervalo de Predição (IP) associado a saída da rede. Medidas de desempenho global normalmente utilizadas para avaliar o desempenho da rede neural, tal como o Erro Médio Quadrático (EMQ), não são capazes de reconhecer regiões onde a resposta da rede possa estar contaminada por incertezas, causadas por fatores como, os erros do modelo devido ao ruído, ou à baixa densidade de dados de treinamentos nestas regiões (PAPADOPOULOS; EDWARDS, 2001). Estas desvantagens ocorrem devido a natureza empírica das redes neurais, onde é difícil perceber quando a rede está extrapolando ou calculando a saída para uma região cujos dados de treinamento eram insuficientes para realizar uma boa aproximação.

Para um dado grau de confiança, o IP é o intervalo na saída do modelo onde uma estimativa do modelo existe com uma dada probabilidade. Incorporar medidas de confiança nas previsões das redes neurais valida a regressão e auxilia os tomadores de decisão a estabelecerem critérios de risco, necessários em muitas aplicações práticas. Portanto, uma estimação confiável e realística do IP é uma área de pesquisa importante e várias abordagens para fornecer medidas de confiança foram desenvolvidas (LEONARD; KRAMER; UNGAR, 1992), (CHRYSSOLOURIS; LEE; RAMSEY, 1996), (VEAUX; SCHWEINSBERG; SHELLINGTON, 1998), (NIX; WEIGEND, 1995), (MCKAY, 1992), (HESKES, 1997) e (NEAL, 1996).

Em geral, as metodologias para cálculo do IP são específicas da arquitetura da rede neural utilizada. Pode-se dizer que existem, basicamente, duas grandes abordagens: a abordagem baseada em aproximação local, onde a estimação do IP se dá através de técnicas de regressão linear; a abordagem global que utiliza modelos de regressão não-linear para cálculo do IP (CHINMAN; DING, 1998).

Abordagens de aprendizagem local, como o modelo de Função de Base Radial (“*Radial Basis Function*” - RBF), possuem como principal característica o conceito de vizinhança. Um ponto, seja de treinamento ou teste, é considerado local a um ponto de teste, quando está espacialmente localizado em uma região

limitada e bem definida em torno deste ponto. Um dos mais conhecidos modelos que usa aprendizagem local é a rede Índice de Validade (“*Validity Index network*” - VInet), que é uma extensão das redes RBF proposta por Leonard et al. (1992) que calcula o IP para sua saída, além de outras medidas de confiança, como a função densidade de probabilidade dos dados e o *flag* de extrapolação.

Por outro lado, modelos de redes baseadas em aprendizagem global, como Perceptrons de Múltiplas Camadas (“*Multilayer Perceptrons*” - MLP), não possuem este conceito de vizinhança local e, portanto, não podem ser facilmente estendidos para incorporar o cálculo do IP. Algumas soluções para calcular o IP em redes MLP foram propostas na literatura (CHRYSSOLOURIS; LEE; RAMSEY, 1996), (SHAO et al., 1997), (VEAUX; SCHWEINSBERG; SHELLINGTON, 1998), (HWANG; DING, 1997), (NEVES; ROISENBERG; NETO, 2009), porém esses algoritmos assumem fortes restrições, tais como: o número de pontos de treinamento devem tender para o infinito (HWANG; DING, 1997); os resíduos devem ser independentes e normalmente distribuídos com média zero (CHRYSSOLOURIS; LEE; RAMSEY, 1996) e a rede deve ser treinada até a convergência (VEAUX; SCHWEINSBERG; SHELLINGTON, 1998). As consequências de violar essas restrições podem se refletir na estimação do IP, que pode ser sub- ou superestimado, perdendo assim sua eficiência como uma medida da incerteza da previsão do modelo.

Devido ao conceito implícito de vizinhança, os modelos locais não necessitam que os resíduos tenham a mesma distribuição normal em todo domínio do problema e pode-se considerar que a variância dos resíduos é constante apenas na região do campo receptivo de uma unidade da RBF, podendo variar de unidade para unidade. Por outro lado, em estudos comparativos foi observado que o tamanho do intervalo de predição, calculado na VInet, nem sempre corresponde à distribuição dos dados de treinamento ou não atingem uma probabilidade de cobertura desejada (SHAO et al., 1997), (YANG et al., 1991). Além disso, a maneira como a densidade é calculada na VInet pode levar a resultados insatisfatórios, principalmente quando se utiliza dados multi-dimensionais (WEDDING; CIOS, 1997).

Com base nessas concepções, a hipótese deste trabalho é que pode-se corrigir as deficiências da VInet e obter medidas de confiança mais satisfatórias e com menos restrições sobre o modelo, que possam, dessa forma ajudar os tomadores de decisão em aplicações reais, como na caracterização de reservatórios de petróleo e gás.

1.2 OBJETIVOS GERAL E ESPECÍFICOS

O objetivo geral deste trabalho é propor uma extensão as técnicas de cálculo de intervalo de predição na VInet, de modo a superar as limitações encontradas no método original proposto por Leonard et al. (1992), tais como incorporar a densidade dos dados de treinamento no intervalo de predição e suportar o cálculo de densidade para dados multidimensionais. Considerando esse objetivo inicial, pretende-se atender os seguintes requisitos:

1. Investigar e identificar as possíveis causas de problemas nas medidas de confiança calculadas pela VInet.
2. Melhorar a resposta do intervalo de predição calculado pela VInet, tornando-o mais compatível com uma probabilidade de cobertura desejada.
3. Oferecer uma alternativa para o cálculo da densidade.
4. Comparar as alterações propostas com o modelo original.
5. Aplicar o método no problema real de caracterização de reservatórios de petróleo.

1.3 JUSTIFICATIVA E RELEVÂNCIA

A VInet, por ser baseada na RBF, é uma rede mais simples que as MLPs, não possui muitas variações na arquitetura e não utiliza nenhum algoritmo de otimização não-linear, o que faz com que seu processo de treinamento seja bastante rápido (BISHOP, 1995). É um modelo que apresenta vários indicadores de confiança, como o IP, a densidade dos dados e o indicador de extrapolação, o que garante mais certeza para o usuário da rede sobre a qualidade da resposta. Em relação à maneira de como as medidas de confiança são calculadas nas MLPs, na VInet este cálculo é feito de uma forma mais intuitiva e com menos restrições.

Por outro lado, apesar de alguns trabalhos apontarem deficiências na VInet, não existem muitas soluções na literatura para corrigi-los (SHAO et al., 1997) (YANG et al., 1991). O presente trabalho torna-se relevante por dar continuidade ao trabalho de Leonard et al. (1992), analisando e estendendo a VInet, de modo a tornar o IP calculado pela rede mais condizente com a probabilidade de cobertura desejada.

1.4 CONTRIBUIÇÕES RELEVANTES

Este trabalho faz as seguintes contribuições técnicas para a área da Inteligência Artificial Conexionista:

- Revisão do estado da arte para o cálculo do intervalo de predição e outras medidas de confianças para redes neurais, principalmente no caso das redes de base radial.
- Proposição de mudanças na arquitetura da rede índice de validade, para que calcule de forma mais precisa as medidas de confiança na saída.
- Aplicação das mudanças em problemas reais de caracterização de reservatório e comparação com outras soluções baseadas em redes neurais.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

A dissertação está organizada da seguinte maneira:

- **Capítulo 2 - Medidas de Confiança para Redes de Função de Base Radial:** Este capítulo descreve os fundamentos sobre o funcionamento das redes de função de base radial. Como ocorre seu treinamento, quais as diferenças para outros modelos de redes e como é calculado o intervalo de predição nesse tipo de rede.
- **Capítulo 3 - Propostas para extensão do modelo VInet:** Este capítulo apresenta as modificações propostas para a rede índice de validade, com o objetivo de melhorar sua aproximação e intervalo de predição na saída.
- **Capítulo 4 - Testes e Resultados:** Este capítulo relata os diferentes experimentos utilizados para validação das modificações propostas por este trabalho. Foram utilizados diversos conjuntos de dados, sintéticos e reais e os resultados comparados com outros modelos de redes neurais.
- **Capítulo 5 - Conclusões:** Este capítulo trata das conclusões e considerações finais sobre o trabalho. Discute as limitações e possíveis direções para trabalhos futuros.

2 MEDIDAS DE CONFIANÇA PARA REDES DE FUNÇÃO DE BASE RADIAL

Uma rede neural é um processador paralelamente distribuído, formado por várias unidades computacionais menores, chamadas de neurônios e que possuem a propriedade de armazenar conhecimento experimental e torná-lo disponível para uso (HAYKIN, 1998). Os neurônios são as unidades fundamentais de uma rede neural e, de acordo com Haykin, seu modelo possui três elementos básicos:

- Um conjunto de sinapses, cada uma possuindo um determinado peso.
- Um somador para somar os sinais de entrada ponderados pelo peso sináptico.
- Uma função de ativação que restringe a amplitude do sinal de saída do neurônio. Geralmente o intervalo de saída é $[0, 1]$ ou $[-1, 1]$.

A Figura 2.1 mostra o esquema dos componentes do neurônio.

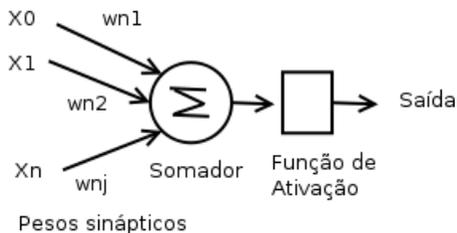


Figura 2.1: Modelo de um neurônio, as entradas são ponderadas por pesos que são somadas pelo somador e depois servem como entrada para a função de ativação.

Dependendo de como os neurônios estão organizados estruturalmente na rede, pode-se classificá-las em: redes de camada única e redes de múltiplas camadas. A forma mais simples de uma rede é a que propaga o sinal no sentido da camada de entrada, diretamente para a camada de saída. Apesar de possuir duas camadas (uma de entrada e outra de saída), essa rede é chamada de rede de única camada, porque apenas a camada de saída possui neurônios com propriedades computacionais, a camada de entrada é apenas uma fonte dos sinais de entrada.

A outra classe de redes neurais, as de múltiplas camadas, apresentam uma ou mais camadas ocultas entre a camada de entrada e a de saída. As unidades de entrada propagam o sinal para a segunda camada (a primeira oculta), os sinais

de saída da segunda camada são, então, utilizados como entrada da camada seguinte, e assim por diante, até a camada final, cujo sinal de saída corresponde a resposta global da rede para o padrão de entrada fornecido pela primeira camada.

As redes de camada única apresentam uma série de limitações em termos de funções que elas podem representar, ao contrário das redes de múltiplas camadas que são capazes de aproximar qualquer função contínua (BISHOP, 1995).

Em relação ao tipo de função de ativação que utilizam nos neurônios, as redes de múltiplas camadas que empregam a função sigmoide são geralmente chamadas de Perceptrons de Múltiplas Camadas (MLP). Neste trabalho, utilizou-se redes que possuem como função de ativação a característica de crescerem (ou decrescerem) monotonicamente, de acordo com a distância para um ponto central. Esse modelo de rede é conhecido como Rede de Função de Base Radial (RBF).

Suponha que se tem um conjunto com Q pares de vetores:

$$(x^1, y^1), (x^2, y^2), \dots, (x^Q, y^Q) \quad (2.1)$$

onde $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^k$. A camada escondida e de a saída possuem respectivamente J e K neurônios. A camada de entrada recebe um sinal x^q e o propaga para a camada escondida. Uma função de ativação frequentemente usada nas unidades escondidas por suas propriedades analíticas bastante conhecidas, é a Gaussiana (ou Normal):

$$f(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (2.2)$$

onde σ é o raio ou abertura da função. Então a saída de um neurônio da camada escondida que utiliza gaussianas como função de ativações é:

$$a_j = \exp(-||x^q - c_j||^2/\sigma_j^2) \quad (2.3)$$

onde c_j é o centro da função de base e σ_j determina até qual distância no espaço de entrada o neurônio terá uma influência significativa. Outras funções de base radial que podem ser utilizadas são a multiquadrática (Equação 2.4), multiquadrática inversa (Equação 2.5) e cauchy (Equação 2.6) (ORR, 1996).

$$f(x) = \sqrt{\sigma^2 + x} \quad (2.4)$$

$$f(x) = \frac{1}{\sqrt{\sigma^2 + x}} \quad (2.5)$$

$$f(x) = \frac{1}{\sigma^2 + x} \quad (2.6)$$

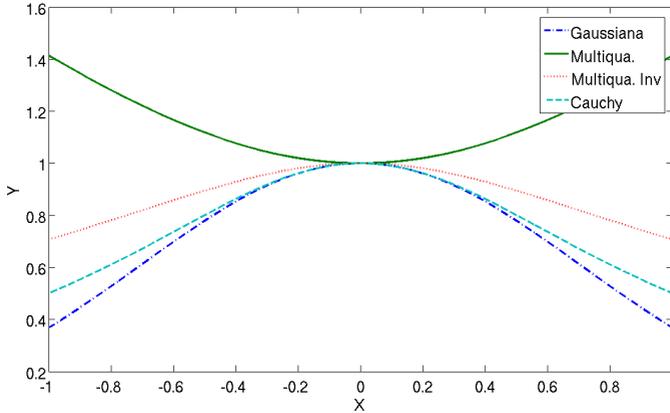


Figura 2.2: Gráfico das funções de base radial mais usadas nas redes RBF, todas foram calculadas com raio 1 e centros 0.

O gráfico dessas funções é ilustrado na Figura 2.2.

A ativação das unidades da camada de saída geralmente é linear, dessa forma, tem-se:

$$a_k = \sum_{j=1}^J w_{kj} a_j + \beta_j \quad (2.7)$$

O termo *bias* pode ser adicionado como uma 'função de base radial extra' fixada sempre em 1. Assim como nas redes MLP, foi provado que a sobreposição linear de várias funções de base também funcionam como aproximadores universais (HARTMAN E. J.; KOWALSKI, 1990) (PARK; SANDBERG, 1991) (PARK; SANDBERG, 1993). E até mesmo possuem vantagens aproximativas que não são observadas nas redes MLP (GIROSI; PODDIO, 1990). A Figura 2.3 mostra a arquitetura de uma rede RBF típica, que apresenta sempre três camadas. A primeira camada de entrada, apenas recebe as informações do ambiente, a segunda camada aplica a transformação não linear do espaço de entrada para o espaço oculto, de maior dimensionalidade e a camada de saída linear fornece a resposta para o sinal de entrada.

2.1 TREINAMENTO DAS REDES RBF

Uma característica importante das redes neurais é sua capacidade de aprender a partir de seu ambiente, através de um processo iterativo de ajustes

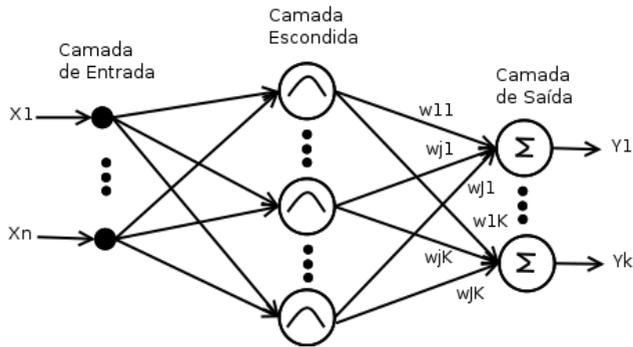


Figura 2.3: Diagrama da arquitetura das redes RBF.

aplicados aos pesos sinápticos. Esse conjunto de passos para solução do problema de aprendizagem é chamado de algoritmo de aprendizagem (HAYKIN, 1998).

As redes RBF têm sua origem ligada às técnicas de interpolação exata em um conjunto de pontos em um espaço multi-dimensional (POWEL, 1987). Essa técnica realiza o mapeamento entre o espaço de entrada e o de saída centralizando uma função de base radial em cada vetor de entrada, dessa forma a função de interpolação passa exatamente através de cada ponto de treinamento. Porém essa metodologia de interpolação exata gera resultados indesejáveis quando existe ruído nos dados, nesse caso a função de interpolação que fornece a melhor generalização deveria ser bem mais suave (BISHOP, 1995).

Através de algumas modificações na técnica de interpolação exata, criou-se o algoritmo de treinamento das redes RBF (MOODY; DARKEN, 1989), (BROMHEAD; LOWE, 1988). A principal mudança foi não utilizar todo o conjunto de treinamento como centros das funções de base, ao invés disso, os parâmetros das funções são determinados pelo algoritmo de treinamento.

O algoritmo de aprendizagem das redes RBF é dividido em dois estágios (MOODY; DARKEN, 1989), (CHEN; MULGREW; MCLAUGHLIN, 1992), (LIPPMANN, 1989). No primeiro são calculados os centros c_j e os raios σ_j das funções de base, utilizando apenas as entradas x^q do conjunto de treinamento. Ignorar os dados de saída em um esquema de treinamento de duas fases como esse pode ser vantajoso em várias aplicações onde dados não rotulados (não classificados) são abundantes e os rotulados são mais escassos. Por exemplo, em alguns casos é fácil conseguir dados não rotulados, mas para obter os rotulados é necessário algum especialista humano, o que limita a obtenção de dados em um certo tempo. Dessa forma, no caso da RBF pode-se usar a grande quantidade de dados não rotulados para otimização dos parâmetros da camada escondida e

utilizar a quantidade menor de dados rotulados para estimar os parâmetros da camada de saída (BISHOP, 1994).

Existem várias estratégias para a seleção dos centros, a mais comum é submeter as entradas do conjunto de treinamento à algum algoritmo de agrupamentos como *k*-médias (DUDA; HART, 1973). Porém o algoritmo tradicional de *k*-médias só pode encontrar uma solução ótima local que depende da escolha inicial dos centros dos agrupamentos (HAYKIN, 1998). Para superar essa limitação Chen (CHEN, 1995) propôs o *algoritmo de agrupamentos de k-médias aperfeiçoado*, que consegue convergir para uma configuração ótima ou próxima da ótima, independentemente das escolhas dos centros iniciais. Também é possível utilizar outros algoritmos neurais de agrupamentos, como por exemplo a rede de Kohonen (KOHONEN, 1990).

Outra abordagem para determinar os centros é a de seleção de subconjuntos, que consiste em escolher um subconjunto dos dados de treinamento como centro das funções de base. Geralmente existem duas maneiras de fazer isso, a *seleção adiante*, onde se começa com apenas um centro e depois é adicionado um novo centro de cada vez, usando como critério a redução máxima do erro na saída. Uma maneira eficiente de fazer seleção adiante é através do *algoritmo de mínimos quadrados ortogonais* (CHEN; BILLINGS; LUO, 1989), (CHEN; COWAN; GRANT, 1991). A outra maneira é a *eliminação inversa*, onde o conjunto inicia com todos os pontos possíveis e então cada um é eliminado por vez, seguindo como critério de remoção, o centro que reduzir mais o erro da rede. Também é possível utilizar os dois princípios de seleção adiante e eliminação inversa ao mesmo tempo, Orr utilizou árvores de regressão para criar uma estrutura hierárquica onde cada nó seria um possível centro da RBF, então o algoritmo percorre a árvore avaliando em cada iteração qual seria a influência de adicionar o nó atual e ainda a influência de se remover os nó pai ou adicionar os dois nós filhos (ORR, 1999).

Quanto aos raios, estes podem ser determinados através da heurística dos vizinhos mais próximos: calcula-se a média da distância euclidiana do centro de um neurônio para os L neurônios mais próximos, onde L é definido pelo projetista. Essa heurística garante a suavidade e continuidade da função aproximada (LEONARD; KRAMER; UNGAR, 1992).

No segundo estágio de treinamento, os parâmetros da camada escondida são fixados e os pesos da camada de saída são determinados através de regressão linear, minimizando-se o erro entre a saída da rede e os valores desejados. É conveniente utilizar como função de erro, a soma dos erros quadráticos:

$$E = 1/2 \sum_{q=1}^Q (a^o(x^q) - y^q)^2 \quad (2.8)$$

onde x^q representa um vetor do conjunto de treinamento com Q vetores, y^q é a saída desejada da entrada x^q e a^o é a saída calculada pela rede para a entrada x^q .

Uma vez que a função de erro é uma função quadrática dos pesos, seu mínimo pode ser obtido através da solução de um conjunto de equações lineares.

$$A^T AW^t = A^T Y \quad (2.9)$$

onde Y é a matriz formada por todas as saídas do conjunto de treinamento, e A é uma matriz onde cada elemento é a saída dos neurônio da camada escondida para cada padrão x^q do conjunto de treinamento, definida por $(A)_{qj} = a_j(x^q)$. A solução formal para os pesos é dada por:

$$W^T = A^{-1}Y \quad (2.10)$$

É possível considerar outra função de ativação para a camada de saída, bem como outra função de erro que não a soma dos erros quadráticos, entretanto a determinação dos pesos não seria mais um problema linear, mas sim um problema de otimização não-linear. Uma das vantagens das redes RBF é justamente o fato de não necessitar de um algoritmo de otimização não-linear.

A seguir são apresentados exemplos para ilustrar da influência que a escolha dos parâmetros das funções de base possui sobre a aproximação calculada pela rede. Nestes exemplos, o mesmo conjunto de treinamento foi criado selecionando-se, aleatoriamente, 30 pontos gerados, usado a equação:

$$y(x) = 0.5\sin(1.5\pi x + \pi/2) + 2.0 \quad (2.11)$$

e adicionado um ruído gaussiano com desvio padrão 0.1. As redes RBF foram criadas com 6 unidades escondidas usando k-médias para selecionar os centros. Os pontos nas figuras representam os dados de treinamento e a linha é a saída calculada pela rede. A Figura 2.4 mostra uma escolha adequada dos raios, baseada na heurística dos vizinhos mais próximos. Já a Figura 2.5 mostra os efeitos de se aplicar um σ muito largo, que causou uma aproximação bem diferente da função que os dados de treinamento representavam. A Figura 2.6 mostra raios bem menores do que deveriam, prejudicando a suavidade da aproximação.

2.2 COMPARAÇÃO COM REDES MLP

Redes RBF e MLP possuem algumas semelhanças como serem redes *feedforward* e aproximadores universais, mas diferem em algumas características:

1. RBFs geralmente possuem três camadas, sempre com unidades escondidas com ativações de base radial seguida por uma camada linear. MLPs

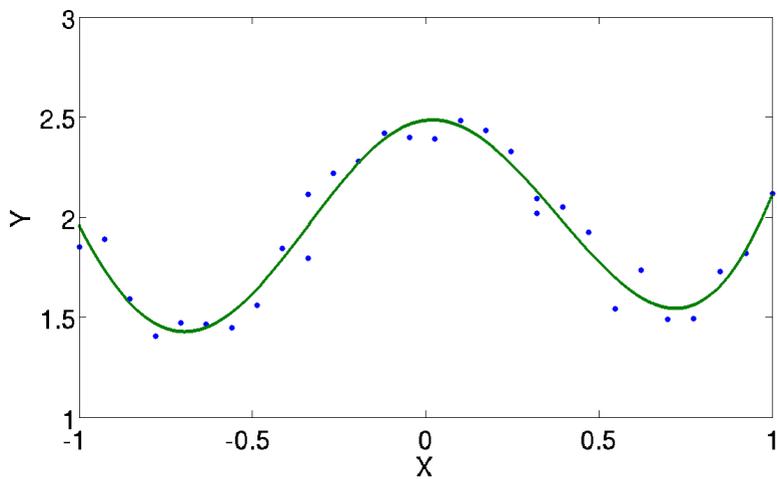


Figura 2.4: Aproximação da rede RBF utilizando raios calculados segundo a heurística dos vizinhos mais próximos.

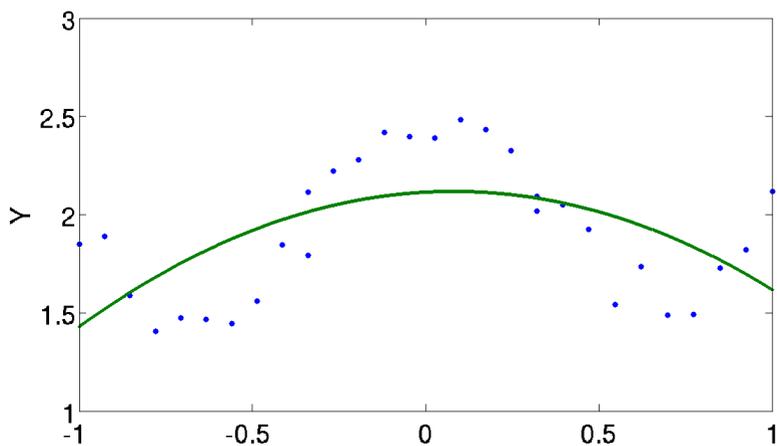


Figura 2.5: Aproximação da rede RBF utilizando raios bem maiores (multiplicados por 1000).

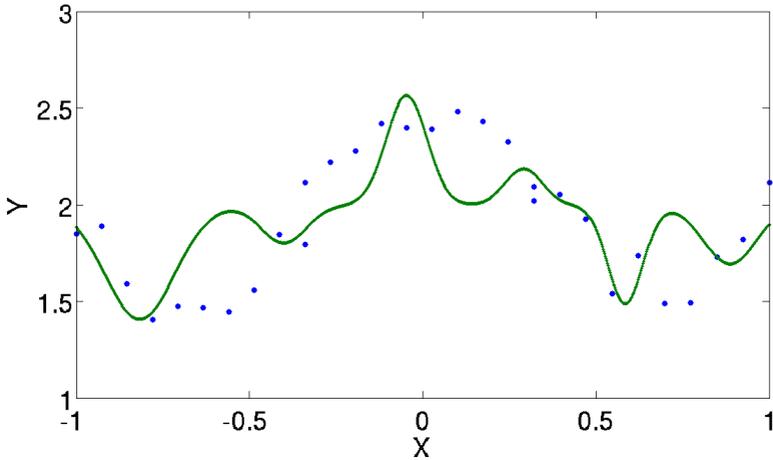


Figura 2.6: Aproximação da rede RBF utilizando raios bem menores (divididos por 5).

podem possuir várias camadas com arquiteturas mais complexas e diferentes funções de ativação para os neurônios.

2. A função de ativação de uma unidade da camada oculta da rede RBF calcula a distância euclidiana entre o vetor e o centro da unidade, enquanto que a MLP calcula o produto interno da entrada pelo peso do neurônio.
3. O treinamento da RBF é feito em duas fases, uma geralmente não supervisionada e outra supervisionada que é determinada através da resolução de um sistema de equações lineares. O que faz com que o treinamento seja bem mais rápido que a otimização não-linear baseado em gradiente, empregada na MLP.
4. Os parâmetros de uma MLP são ajustadas utilizando um esquema de treinamento global e supervisionado, cujo objetivo é minimizar alguma medida de erro global do modelo. As RBFs utilizam funções não-lineares localizadas com decaimento exponencial, empregando dessa forma, o conceito de vizinhança, e criando uma aproximação local para o mapeamento entrada-saída. Isso faz com que uma MLP necessite de um número menor de parâmetros que uma rede RBF para realizar um mapeamento entrada-saída com a mesma precisão. (HAYKIN, 1998)

2.3 REDE ÍNDICE DE VALIDADE

A VInet é uma extensão da rede RBF que calcula o intervalo de predição (IP) e outras medidas de confiança na saída (LEONARD; KRAMER; UNGAR, 1992). Seu uso já foi aplicado em casos reais, como por exemplo em problemas de controle (VOUMVOULAKI; HATZIARGYRIOU, 2010), (TSAI JI-ZHENG CHU; SHIEH, 2003), (MUNOZ; SANZ-BOBI, 1998), na previsão de séries temporais (WEDDING-II; CIOS, 1996) e outros.

A VInet tem dois tipos de medidas de confiança, um deles indica se a rede está ou não extrapolando, caso esteja então a saída da rede é imprecisa, pois a rede não recebeu dados de treinamento suficientes de uma certa região de entrada. Caso a rede não esteja extrapolando, então utiliza-se o outro tipo de medida de confiança que indica o quão boa está a aproximação, essa medida é exibida como intervalos de predição (LEONARD; KRAMER; UNGAR, 1992).

Uma das medidas é o indicador de extrapolação, max-act, que é a maior ativação das unidades escondidas para um dado de teste x :

$$\text{max-act} = \max \{a_j(x)\} \quad (2.12)$$

Como cada unidade escondida define uma região centrada em um ponto, um baixo valor de max-act significa que o ponto de teste está distante de todos os centros dessas unidades e consequentemente existem poucos dados de treinamento nessa vizinhança do ponto de teste. Embora essa medida indique uma possível extrapolação, ela não define a quantidade ou a densidade dos pontos de treinamento, apenas identifica que existe "algum" dado de treinamento na região. Também pode ocorrer de valores baixos serem obtidos entre regiões de interpolação (LEONARD; KRAMER; UNGAR, 1992).

Uma medida mais satisfatória para caracterizar a extrapolação é encontrar a função densidade de probabilidade. Uma maneira de fazer isso é através do procedimento conhecido como Janelas de Parzen (PARZEN, 1962). Este algoritmo é classificado como não paramétrico, pois não assume a priori nenhuma 'forma' da função densidade de probabilidade que se deseja modelar. A estimativa é feita sobrepondo várias funções de *kernel* em cada padrão observado, dessa forma, cada observação contribui com a estimativa. Suponha que se deseja determinar a densidade $p(x)$ em um ponto x , então coloca-se uma função janela no ponto x , e determina-se quantas observações x_i caem dentro da janela. A $p(x)$ final é a soma de todas as contribuições que caíram nessa janela. Existem algumas maneiras de implementar Janelas de Parzen em redes neurais, em uma delas utilizou-se a mesma quantidade de unidades escondidas que o número de padrões observados (SPECHT, 1990), em outra abordagem os pesos da rede, os centros e os raios foram otimizados para a tarefa de estimação de densidade ao

invés de regressão (TRAVEN, 1991).

O procedimento proposto por Leonard et al. (1992) na VInet é diferente e é feito em duas partes, primeiro a densidade é calculada para cada unidade escondida, durante o treinamento da rede:

$$p_j = \frac{\sum_{q=1}^Q a_j(x^q)}{Q(\pi^{1/2}\sigma_j)^D} \quad (2.13)$$

onde D é o número de dimensões dos dados. Depois, baseado nessas densidades locais, a densidade para um ponto qualquer x é calculado através de uma média ponderada:

$$p(x) = \frac{\sum_{j=1}^J a_j(x)p_j}{\sum_{j=1}^J a_j(x)} \quad (2.14)$$

onde as ativações a_j funcionam como função de pertinência do ponto x para as unidades escondidas; p_j é a densidade local de cada unidade escondida que são combinadas na proporção da ativação para o ponto de teste. Se este ponto está próximo de um dos centros, então a ativação será alta e a densidade local correspondente a esta unidade irá aparecer na fórmula com um peso próximo de 1. Caso o ponto de teste esteja longe, então o peso será menor e a contribuição dessa unidade será mais baixa.

Para interpretar o significado de $p(x)$ e saber se a quantidade de dados foi suficiente, é necessário estimar um valor de limiar. Uma heurística para essa escolha é utilizar o menor $p(x^q)$ do conjunto de treinamento, e caso seja necessário o projetista pode aumentar ou diminuir este número (LEONARD; KRAMER; UNGAR, 1992).

O cálculo do IP acontece de maneira semelhante à densidade, primeiro é calculado um IP local para cada unidade escondida durante a fase de treinamento e, depois, é feita uma média dos limites de predição de todos os neurônios da camada oculta ponderados pela contribuição de cada unidade escondida.

Assumindo-se que os pontos de treinamento são associados à cada unidade escondida na mesma proporção de suas ativações, a estimação da variância dos resíduos para cada unidade escondida j é dada por:

$$S_j^2 = \frac{\sum_{q=1}^Q a_j(x^q)(y^q - f(x^q; \hat{\theta}))^2}{n_j - 1} \quad (2.15)$$

onde $(x^q - f(x^q; \hat{\theta}))$ é a diferença entre o valor desejado y^q e $f(x^q; \hat{\theta})$ é o valor obtido pela rede, para o dado de treinamento x^q . A ativação tem o papel de função de pertinência do dado de treinamento x^q para a unidade escondida j ,

dessa forma cada ponto de treinamento contribui um pouco para a variância na proporção da sua ativação.

n_j é o número de pontos de treinamento associados em cada unidade escondida j :

$$n_j = \sum_{q=1}^Q a_j(x^q) \quad (2.16)$$

O IP para 95% de confiança é dado por:

$$IP_j = t_{n_j-1}^{\alpha/2} S_j \left(1 + \frac{1}{n_j} \right)^{1/2} \quad (2.17)$$

onde $t^{\alpha/2}$ é o valor da distribuição de t de Student para 95% de confiança com $n^h - 1$ graus de liberdade. Finalmente o IP para um ponto de teste é a média dos IPs locais, ponderados pela contribuição de cada unidade escondida.

$$IP(x) = \frac{\sum_{j=1}^J a_j(x) IP_j}{\sum_{j=1}^J a_j(x)} \quad (2.18)$$

onde a_k^h é a saída para a k -ésima unidade escondida. Para pontos de teste longe dos centros dos neurônios escondidos, todas as ativações vão ser baixas. Através da divisão pela soma das ativações, o valor do IP se aproxima do valor do IP^h da unidade escondida mais próxima ao ponto de teste (LEONARD; KRAMER; UNGAR, 1992).

Este método assume que os resíduos do modelo são independentes e normalmente distribuídos com média zero e variância constante dentro da região definida por cada unidade escondida, mas variando de unidade para unidade.

A Figure 2.7 mostra a arquitetura da rede VInet. Para simplificar o diagrama, a rede foi representada com apenas uma entrada, dois neurônios na camada escondida e uma na de saída. Os cálculos das medidas de confiança são realizados por neurônios extras que estendem o modelo d RBF comum.

Para resumir todo o processo para usar a VInet:

- Encontre os valores dos raios σ_j e dos centros c_j das unidades escondidas.
- Calcule os pesos da camada de saída.
- Armazene os erros durante o treinamento.
- Calcule S_j^2 e todas as outras medidas locais como n_j , p_j , e IP_j .
- Quando um novo dado é apresentado a rede, calcule a saída, o max-act, IP final e densidade final.

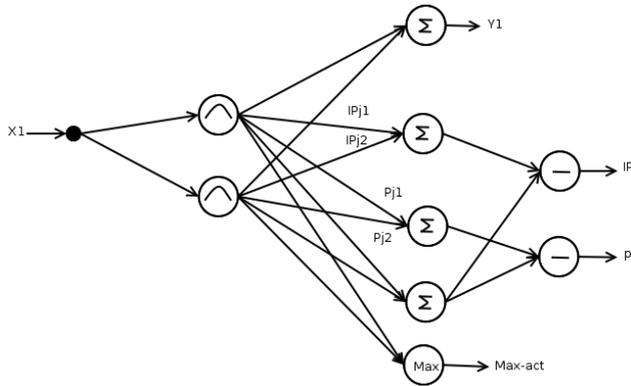


Figura 2.7: Arquitetura da rede VInet.

2.4 LIMITAÇÕES DO INTERVALO DE PREDIÇÃO DA VINET

A confiabilidade na resposta da rede depende de três fatores. Primeiro depende da modelagem da rede, se existe um número suficiente de pesos, ajuste adequado dos parâmetros do algoritmo de aprendizagem, dentre outros. Segundo fator importante é densidade dos dados de treinamento, quando um novo dado é apresentado a rede, caso seja pertencente à uma região do espaço de entrada com alta densidade, então a rede estará interpolando o resultado com base nos dados de treinamento e, conseqüentemente, a resposta será melhor. Por outro lado, se o dado é de uma região de baixa densidade, então esse valor será uma novidade para a rede e existe uma possibilidade do valor previsto pela rede ser diferente do valor real do sistema (BISHOP, 1994), (SHAO et al., 1997). O terceiro e último fator é em relação ao ruído presente nos dados, que é responsável por causar uma variância ao redor da média dos valores desejados. A VInet é capaz de identificar variações de ruído no espaço de entrada, devido suas características de aproximador local, cada unidade escondida calcula sua variância e IP local, que depois são combinados para encontrar o IP final. Dessa forma, caso existam regiões com diferentes níveis de ruído, a RBF consegue ajustar seu IP tornando-o mais compatível com a variância local de cada unidade escondida.

O primeiro fator, pode ser tratado através de técnicas de validação cruzada, onde tenta-se obter os melhores parâmetros do modelo da rede. Então não será discutido neste trabalho as conseqüências para o IP de uma modelagem errada. A influência do ruído no IP também já é bem capturada pela VInet, de forma que o resto desse capítulo se concentra em investigar como a densidade dos dados de treinamento influenciam o IP.

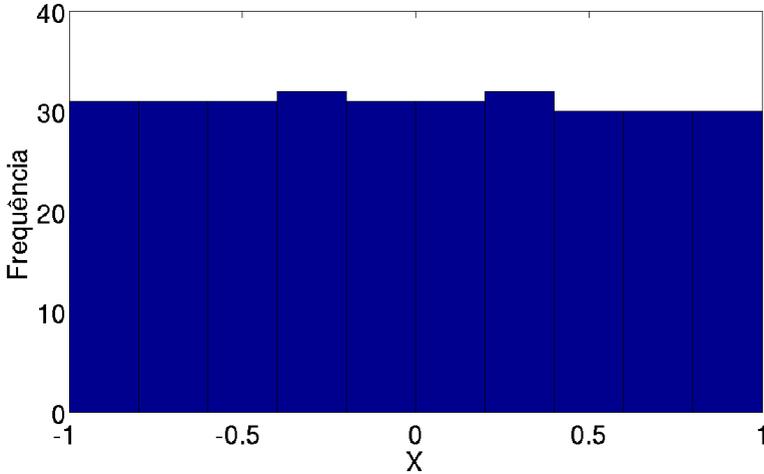


Figura 2.8: Histograma dos dados de treinamento com densidade uniforme.

2.4.1 Influência da Densidade no Intervalo de Predição

Segundo (YANG et al., 1991) os intervalos de predição da VInet refletem pouco a densidade dos dados de treinamento. Isso acontece principalmente nas regiões de extrapolação, onde a rede teve poucos ou nenhum dado de treinamento. Para ilustrar esse caso, treinou-se uma rede com 310 dados contidos no intervalo $[-1, 1]$, gerados a partir da equação 2.11. A densidade dos dados é uniforme, como mostra o histograma na figura 2.8. A saída da rede foi calculada para pontos no intervalo entre $[-1, 2]$. A Figura 2.9 mostra a saída da rede, os intervalos de predição criados com 95% de confiança, e dados de teste. Não existe nenhum dado de treinamento no intervalo $[1, 2]$ do eixo x , então a rede está extrapolando todos esses valores e calculando uma saída errada. Porém nenhuma mudança visível ocorre nos intervalos de predição, a Figura 2.10 mostra apenas o IP para cada ponto de teste, percebe-se que no intervalo $[1, 2]$ ele fica estável ao invés de aumentar.

Quais as consequências do IP não refletir a extrapolação? Talvez a mais importante é do IP acabar falhando na sua função de mensurar a incerteza da rede e acaba prejudicando os usuários que confiam que o IP está bom por estar com uma largura aceitável quando na realidade a rede realiza uma aproximação completamente equivocada. Além disso o IP não apresenta um nível de confiança desejado. Uma maneira de medir a eficiência dos IP é calculando a probabilidade de cobertura do intervalo de predição (PCIP), que consiste na porcentagem de

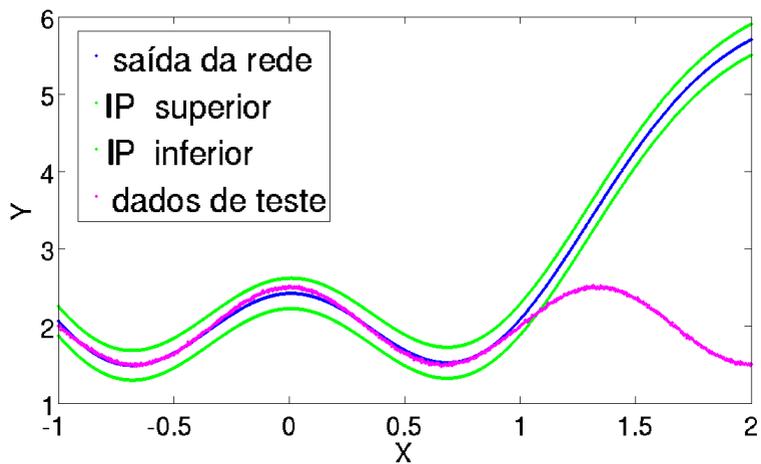


Figura 2.9: Saída da rede, intervalos de confiança e saída desejada.

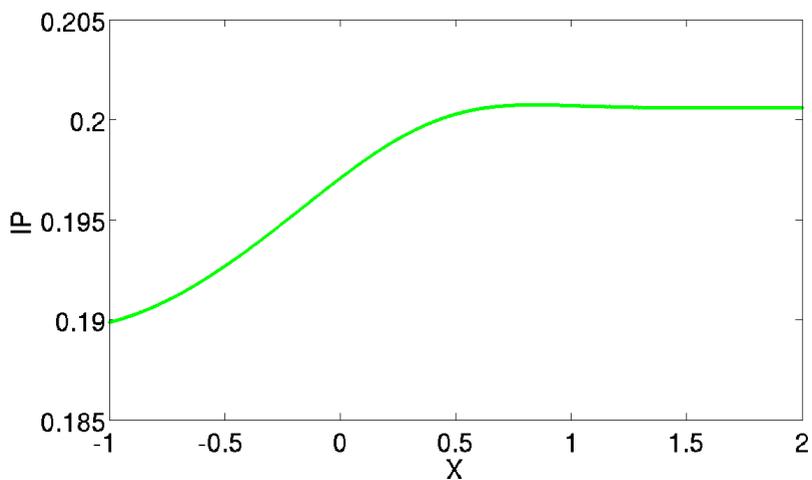


Figura 2.10: Intervalo de predição para os dados de teste.

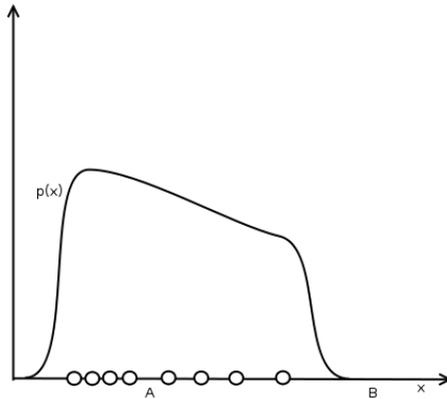


Figura 2.11: Esquema ilustrando como se utiliza a densidade como medida de confiança.

pontos de teste que caem dentro da região definida pelo IP (PAPADOPOULOS; EDWARDS, 2001). Quando o intervalo não reflete a incerteza na estimação, o PCIP fica abaixo do nível que seria esperado e mais uma vez o tomador de decisão pode não se sentir seguro na sua escolha. Na próxima seção é detalhado como a densidade é calculada na VInet, já que ela pode ser utilizada como um outro indicador de confiança que identifica extrapolação. Depois são investigadas as vantagens e desvantagens dessa maneira de calcular densidade e como podemos alterar o IP para exibir o PCIP adequado.

2.4.2 Cálculo da Densidade na VInet

Uma das medidas de confiança fornecida pela VInet é a densidade, o procedimento para calculá-la consiste em usar os dados de treinamento para estimar a densidade $\hat{p}(x)$ da densidade desconhecida $p(x)$. Quando a rede está pronta para ser utilizada, cada entrada apresentada à rede também é usada para calcular $\hat{p}(x)$. Assim um valor muito baixo de densidade indica que a rede não é familiar com esse dado e que ele é muito diferente dos padrões usados no conjunto de treinamento, sendo dessa forma, uma novidade e passível de gerar uma saída com erro elevado.

A Figura 2.11 mostra como seria a utilização da densidade por um usuário da rede. Os círculos no eixo x representam os dados de treinamento e com base neles é estimada a densidade $\hat{p}(x)$. Para um ponto de teste A o valor de $\hat{p}(x)$ é maior, indicando similaridade com os dados do conjunto de treinamento, já o ponto B é de uma região com menor valor de $\hat{p}(x)$ e, então, é considerado uma novidade para a rede.

Na VInet, a densidade é calculada através de Janela de Parzen, encontrando uma densidade local para cada unidade escondida e depois combinando todas elas para determinar a densidade final. Um problema dessa abordagem é que nem sempre o número ótimo de unidades necessárias para calcular a densidade é o mesmo número ótimo de unidades escondidas necessárias para aproximação da rede. Ainda mais porque o resultado da Janela de Parzen melhora quando todos os pontos são usados para a estimação.

Para ilustrar esse problema, a figura 2.12 mostra o histograma de 310 dados de treinamento contidos no intervalo $[-1, 1]$ gerados com a equação 2.11. Os dados estão divididos em 3 regiões distintas, com densidade maior a medida que o valor no eixo x aumenta. Duas redes foram criadas e treinadas com esse mesmo conjunto, uma delas com 4 neurônios na camada escondida e outra utilizando 310 neurônios, onde cada dado do conjunto de treinamento era o centro de uma unidade escondida. A Figura 2.13 mostra a densidade calculada com a rede de 310 unidades escondidas e a 2.14 mostra a densidade calculada pela rede com menos unidades. Percebe-se que a rede com mais unidades obteve um resultado muito mais próximo ao histograma, enquanto que a densidade da rede com menos unidades ficou bastante prejudicada. Por outro lado, a Figura 2.15 mostra como ficou a aproximação da rede que utilizava 310 unidades e a Figura 2.16 mostra a aproximação para a rede que possuía apenas 4 unidades. A rede com menos unidades consegue aproximar melhor a função enquanto que a rede com mais unidades acaba sofrendo de *overfitting* e aprendendo o ruído.

Outra desvantagem do cálculo da densidade na VInet é que para dados com muitas dimensões, o valor da densidade fica muito pequeno, devido ao expoente no denominador na equação 2.13, a ponto de resultar em problemas de arredondamento e tenderem a zero (WEDDING; CIOS, 1997). Wedding et al. mostra vários exemplos da utilização da VInet em problemas de dimensionalidade alta, em um deles utiliza um banco de dados de um problema real de classificação com 208 entradas de 60 dimensões. Nesse exemplo o denominador na equação 2.13 ficou na ordem de 10^{19} e os valores do p^h foram arredondados para 0, resultando numa densidade final $p(x)$ também 0. Isso significa que não tem como determinar quando os resultados estão ou não extrapolando, perdendo o sentido na utilização desta medida de confiança.

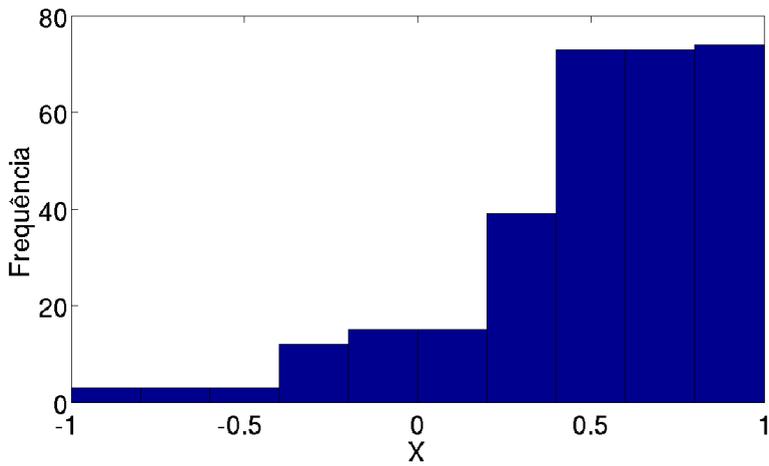


Figura 2.12: Histograma dos dados de treinamento de dados não uniformes.

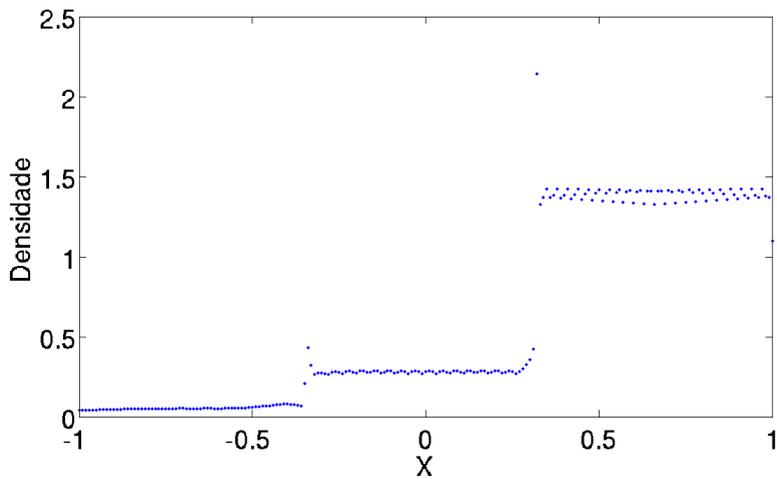


Figura 2.13: Densidade calculada pela rede com 310 unidades escondidas.

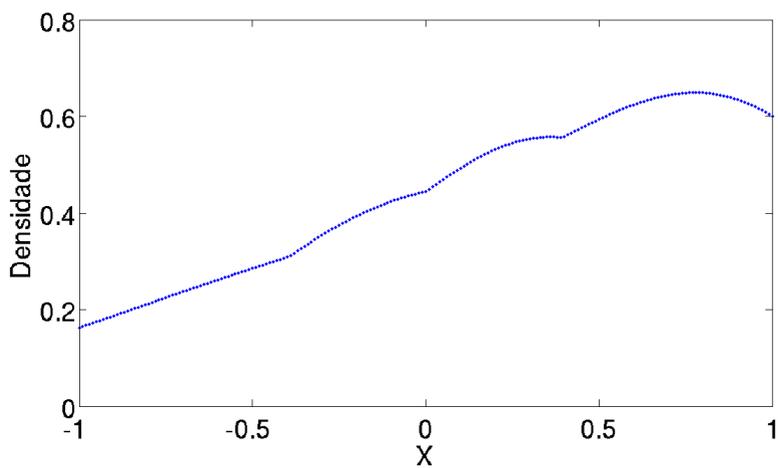


Figura 2.14: Densidade calculada pela rede com 4 unidades escondidas.

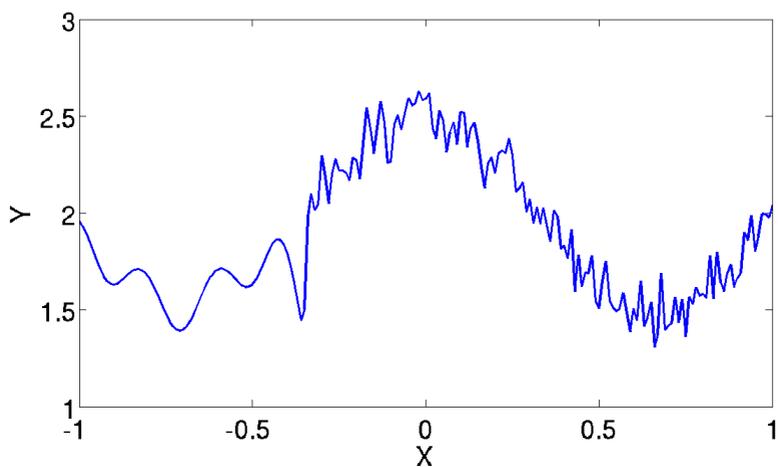


Figura 2.15: Saída da rede quando utiliza-se 310 unidades escondidas.

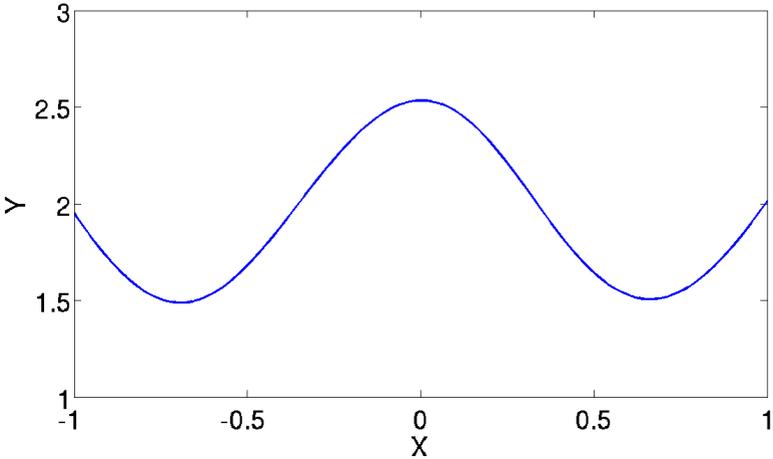


Figura 2.16: Saída da rede quando utiliza-se 4 unidades escondidas.

3 PROPOSTA PARA EXTENSÃO DO MODELO VINET

No capítulo anterior foram apresentadas limitações nas medidas de confiança da VInet, como:

- O intervalo de predição não reflete a densidade dos dados e não garante o PCIP desejado.
- A mesma rede otimizada para realizar uma boa aproximação não garante um resultado realístico de densidade.
- Utilizar Janelas de Parzen pode causar problemas de arredondamento em dados com muitas dimensões.

A solução proposta utiliza um modelo neural bem conhecido, descrito neste capítulo, para calcular a densidade, sem levar em conta as limitações de usar Janelas de Parzen. Além disso, o cálculo do IP foi alterado para incorporar a densidade e obter um PCIP mais próximo do esperado.

3.1 CALCULANDO DENSIDADE COM SOM

Mapas auto-organizáveis são uma classe especial de redes neurais onde os neurônios são organizados em uma grade uni- ou bidimensional. Essa grade passa um processo de aprendizagem competitiva, onde os neurônios competem para serem ativados e os que ganham a competição são chamados de neurônios vencedores. Após o treinamento, a grade passa a se tornar seletivamente sintonizada a vários padrões de entrada (KOHONEN, 1990). Uma propriedade resultante desse processo de aprendizagem é a formação de um mapa topográfico dos dados de treinamento nos quais as coordenadas dos neurônios da grade refletem características intrínsecas dos dados (HAYKIN, 1998). Devido a essa propriedade, as redes SOM são largamente empregadas em problemas de análise de agrupamentos e visualização de dados multivariados.

Para descrever o processo de aprendizagem da rede SOM, considere X o espaço contínuo de dados de entrada onde se conhece um conjunto finito de amostras, ou dados de treinamento, formado por vetores $x \in X$ com dimensão Q tal que $x = [x^1, x^2, \dots, x^q]$. A topologia do espaço de saída A é formado por um arranjo geométrico de neurônios $i \in A$, então o mapeamento do espaço de entrada X para o espaço de saída A é realizado através da transformação não-linear:

$$\Phi : X \rightarrow A \tag{3.1}$$

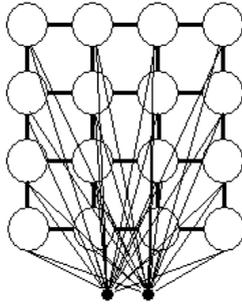


Figura 3.1: Elementos básicos de uma rede SOM, os l neurônios da grade estão dispostos em um formato retangular, no instante n , um padrão x é apresentado a rede e dentre todos os neurônios é escolhido o vencedor v . A vizinhança é definida por h_{vi} centrada no vencedor e apenas esses neurônios possuem suas conexões ajustadas de forma que se aproximem do padrão de entrada.

Cada neurônio na grade possui um vetor de pesos sinápticos com a mesma dimensão dos dados de entrada, representado por $w_i = [w_{i1}, w_{i2}, \dots, w_{iq}]$ onde i é índice do neurônio da grade formada por um total de l neurônios. Os neurônios do mapa estão organizados espacialmente em uma configuração geralmente retangular ou hexagonal e a saída da rede é o índice do vencedor. A Figura 3.1 mostra um exemplo de uma rede SOM com uma grade formada por 16 neurônios em configuração retangular e com uma entrada de tamanho 2.

O algoritmo de treinamento e formação do mapa é não-supervisionado e auto-organizado. O primeiro passo consiste em atribuir valores pequenos aos pesos sinápticos da grade, dessa forma nenhuma organização prévia é imposta à rede. Após essa inicialização, o treinamento ocorre em 3 processos distintos: cooperação, competição e adaptação. Os passos do algoritmo são resumidos nas operações abaixo:

- **Amostragem:** retire uma amostra x para ser aplicada à grade.
- **Casamento por similaridade:** encontre o neurônio vencedor $i(x)$ na iteração n usando o critério de mínima distância euclidiana: $i(x) = \operatorname{argmin}_j \|x(n) - w_j\|$, $j = 1, 2, \dots, l$
- **Atualização:** Ajuste os vetores de peso sináptico de todos os neurônios usando a fórmula $w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n))$ onde $\eta(n)$ é a taxa de aprendizagem e $h_{j,i(x)}(n)$ é a função de vizinhança do vencedor $i(x)$. Ambos parâmetros variam dinamicamente durante a aprendizagem.

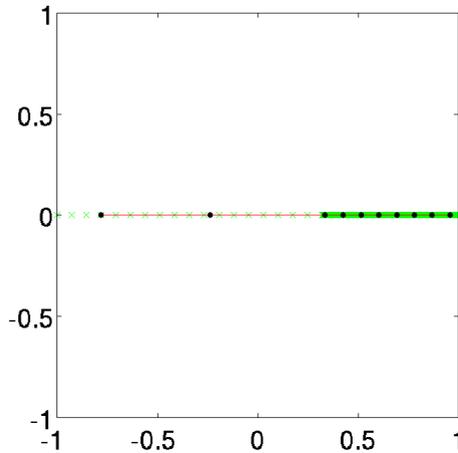


Figura 3.2: Posição dos pesos dos neurônios de uma SOM treinada com dados não uniformemente distribuídos. Os neurônios são representados pelos círculos e os pontos de treinamento pelos caractere x.

- *Continuação*: Continue com o primeiro passo até que não sejam observadas modificações significativas no mapa.

Mais informações sobre as propriedades e processo de aprendizagem das redes SOM pode ser obtido em (HAYKIN, 1998).

Uma característica importante da SOM é o *casamento de densidade*. O mapa de características reflete as variações da distribuição de entrada: regiões do espaço de entrada X de onde os vetores de treinamento x são retirados com maior frequência são mapeados para domínios maiores do espaço de saída A . Por outro lado, as regiões em X das quais são retirados vetores x com baixa probabilidade de ocorrência, serão representados por regiões menores do espaço de saída de A .(HAYKIN, 1998), (HOLMSTROM; HAMALAINEN, 1993), (LUTTRELL, 1991) , (RITTER, 1991).

A Figura 3.2 ilustra essa característica da SOM, para um mapa unidimensional com 10 neurônios. Existem 10 vezes mais pontos de treinamento na região $[0.3, 1]$ do que no resto, então os pesos neurônios tenderam a se concentrar mais nesse local, por outro lado todo resto do espaço de entrada ficou representado por apenas dois neurônios.

Considerando essa característica intrínseca do casamento de densidade da SOM, considerou-se que ela seria adequada para calcular a densidade para

problemas de alta dimensionalidade. Uma das maneiras de estimar a densidade utilizando a SOM é utilizando o conceito de *mistura de modelos*, que representa a distribuição por meio de uma combinação linear de algumas funções de *kernel* (BISHOP, 1995).

A densidade de um dado x é dada por:

$$p(x) = \sum_{j=1}^m P(j)p(x|j) \quad (3.2)$$

onde m é o número de neurônios da SOM, $P(j)$ são os pesos da mistura e podem ser interpretados como a probabilidade *a priori* do dado pertencer a célula do diagrama de Voronói correspondente ao neurônio j e $p(x|j)$ representa a densidade condicional do vetor desejado x para o j -ésimo *kernel*.

As probabilidades $P(j)$ devem satisfazer as condições

$$\sum_{j=1}^m P(j) = 1 \quad (3.3)$$

$$0 \leq P(j) \leq 1 \quad (3.4)$$

Similarmente, as probabilidades $p(x|j)$ devem ser normalizadas para garantir que

$$\int p(x|j)dx = 1 \quad (3.5)$$

Para as funções de *kernel*, uma escolha comum é a gaussiana, onde o centro é o peso do neurônio e o raio é o campo de ativação desse neurônio.

A versão completa do algoritmo é:

- Treine a SOM e utilize os pesos de cada neurônio do mapa como os centros das funções de *kernel*.
- Calcule as probabilidades $P(j) = \frac{N_j}{N}$ onde N_j é o número de dados que caem na célula do diagrama de Voronói de cada neurônio j .
- Calcule os raios de cada função de *kernel*.

Para encontrar o valor dos raios, uma maneira é usar a distância euclidiana do centro de um neurônio j para os k -ésimos centros mais próximos (TERRELL; SCOTT, 1992).

3.2 COMBINANDO DENSIDADE COM INTERVALO DE PREDIÇÃO

A distribuição dos dados de treinamento exerce uma forte influência na confiabilidade da predição do modelo (SHAO et al., 1997). A confiança na predição da rede é maior em regiões onde existem maior quantidade de dados de treinamento, ou seja o IP deveria ser mais estreito. Por outro lado a predição deve ser menos confiável nos locais onde os dados de treinamento são escassos e consequentemente o IP deve ser mais largo.

Uma sugestão dada por Shao et al. (1997) é alterar a fórmula do IP para uma dada entrada x para adicionar a influência da densidade no IP:

$$a^o + IP\beta \quad (3.6)$$

onde a^o é a saída da rede, IP é o intervalo de predição e β é o coeficiente na proporção inversa da densidade dos dados de treinamento e pode ser escrito como:

$$\beta = \frac{a}{b + p(x)} \quad (3.7)$$

onde a e b são constantes desconhecidas e $p(x)$ é a densidade calculada pela SOM para um ponto x . Através de resultados de simulações Shao et al. afirmam que o IP é duas vezes maior em regiões onde os dados não estão disponíveis, então $\beta \simeq 2$ quando $p(x) = 0$. Por outro lado, nas regiões onde $p(x)$ atinge o valor máximo p_{max} , a densidade é suficientemente densa e o IP deve ser igual ao seu valor normal, então:

$$\beta = \frac{a}{b + p(x)} = 1 \quad (3.8)$$

quando $p(x) = p_{max}$. Então, com $a = 2p_{max}$ e $b = p_{max}$ o IP passa a ser calculado como:

$$IP_f(x) = \frac{2IP(x)}{1 + \frac{\hat{p}(x)}{p_{max}}} \quad (3.9)$$

Dessa forma, regiões com densidade próxima a zero vão ter o IP com o dobro do tamanho que um dado de uma região com densidade próxima a p_{max} .

4 TESTES E RESULTADOS

Neste capítulo são apresentados testes mostrando como se comporta o intervalo de predição da VInet sob diversas configurações de densidade e ruído. O primeiro conjunto de testes utiliza apenas dados artificiais controlados, o segundo é um problema mais difícil e multidimensional e o terceiro é uma aplicação real de caracterização de reservatórios de petróleo. Além de utilizar redes RBF, a fim de comparação também será avaliado o IP calculado pelas redes MLPs.

O objetivo desses testes são:

- Demonstrar a influência de diferentes configurações de ruído e densidade no IP.
- Mostrar que a inclusão da densidade no IP discutida no capítulo anterior fornece IPs que refletem melhor a incerteza na aproximação.

Primeiramente é discutida a metodologia empregada nos experimentos e, em seguida, os testes e a análise dos resultados.

4.1 METODOLOGIA

Em uma tarefa real, geralmente interessa apenas a previsão da saída. Neste trabalho, o objetivo é avaliar a performance dos intervalos de predição, comparando a probabilidade de cobertura (PCIP) observado com o nível de confiança do intervalo criado. Uma maneira de calcular o PCIP observado é encontrar a porcentagem de pontos de teste que caem dentro dos intervalos de predição (PAPADOPOULOS; EDWARDS, 2001), (DONALDSON; SCHNABEL, 1987), (HWANG; DING, 1997), (SHRESTHA; SOLOMATINE, 2006).

Para cada experimento foram separados dois conjuntos de dados, um de treinamento e um de testes. Foram criadas 100 instâncias de redes e, depois, calculado o PCIP médio, para que a variação nos pesos não influenciem significativamente na avaliação dos métodos. Cada uma foi treinada com os mesmos parâmetros (taxa de aprendizagem e número de unidades escondidas) e, depois, para cada rede foi apresentado o conjunto de testes, calculado o IP e, então, estimado o PCIP para todo o conjunto.

Além de testes comparativos entre a VInet e as mudanças propostas neste trabalho, também foi avaliado o desempenho das redes MLP, cuja descrição da metodologia para estimação do IP pode ser encontrada em (CHRYSSOLOURIS; LEE; RAMSEY, 1996) e o Método Delta Estendido (MDE) (NEVES; ROISENBERG; NETO, 2009), que incorpora uma ideia semelhante a da VInet nas re-

des MLP: a variância dos ruídos é calculada para diferentes regiões do espaço de entrada, garantindo assim, ao contrário da MLP, um IP que reflete possíveis variações do ruído nos dados. Mais informações sobre o MDE podem ser encontradas em (NEVES, 2009).

Apenas para facilitar a exibição dos resultados nas tabelas, a VInet como apresentada no artigo original foi chamada de S-VInet (Standard VInet), a VInet com densidade incorporada no IP é chamada de E-VInet (Extended VInet), a MLP de S-MLP (Standard MLP) e o MDE de E-MLP(Extended MLP).

4.2 EXPERIMENTO 1

Para esse primeiro teste a equação 2.11 foi usada para criar diferentes conjuntos de dados. Os dados foram gerados com os valores de entrada x pertencente ao intervalo $[-1, 1]$, mas com diferentes configurações de ruído e densidade:

- Conjuntos 1 e 2 possuem três regiões com diferentes níveis de ruído. Para $-1 \leq x < -0.3\sigma_v = 0.05$. Para $-0.3 \leq x < 0.3\sigma_v = 0.1$. Para $0.3 \leq x \leq 1\sigma_v = 0.4$. Conjunto 1 possui 100 pontos de treinamento e o Conjunto 2 possui 1000 pontos.
- Conjuntos 3 e 4 possuem a mesma distribuição de ruído que os Conjuntos 1 e 2, e uma correlação positiva entre o ruído e a densidade, ou seja, as regiões com mais ruído são as que possuem mais dados.
- Conjuntos 5 e 6 possuem a mesma distribuição de ruído que os Conjuntos 1 e 2, e uma correlação negativa entre ruído e densidade. As regiões com mais ruído são as com menor densidade.
- Conjunto 7 possui uma distribuição não uniforme da densidade mas o ruído v possui um desvio padrão de $\sigma_v = 0.1$

Todos os conjuntos estão resumidos na Tabela 4.1 e Tabela 4.2.

O conjunto de testes foi criado com 1000 pontos, gerados com a mesma equação e configuração de ruído, porém com densidade uniforme. Todas as redes foram criadas com 5 unidades escondidas e os intervalos de predição calculados para 95% de confiança. A SOM utilizada para calcular a densidade na E-VInet possuía 10 neurônios.

A Tabela 4.3 apresenta o PCIP médio para todos os testes. Um mau caso para a S-MLP foi o Conjunto 1 e 2 onde existem diferentes níveis de ruído para diferentes regiões do espaço de entrada, por outro lado S-VInet e E-VInet obtiveram resultados melhores devido a abordagem de aproximação local que os

Tabela 4.1: Resumo da densidade e ruído nos dados usados no primeiro teste. A variação da densidade significa como o conjunto de treinamento está distribuído em cada um dos três intervalos do espaço de entrada. A fração é a razão dos dados que pertence a cada intervalo.

Conjunto	Ruído σ_v	Densidade
1	$-1 \leq x < -0.3, \sigma_v = 0.05$	uniforme
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	
2	$-1 \leq x < -0.3, \sigma_v = 0.05$	uniforme
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	
3	$-1 \leq x < -0.3, \sigma_v = 0.05$	$-1 \leq x < -0.3, \rho = 1/6$
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	$-0.3 \leq x < 0.3, \rho = 2/6$
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	$0.3 \leq x \leq 1, \rho = 3/6$
4	$-1 \leq x < -0.3, \sigma_v = 0.05$	$-1 \leq x < -0.3, \rho = 1/6$
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	$-0.3 \leq x < 0.3, \rho = 2/6$
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	$0.3 \leq x \leq 1, \rho = 3/6$
5	$-1 \leq x < -0.3, \sigma_v = 0.05$	$-1 \leq x < -0.3, \rho = 3/6$
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	$-0.3 \leq x < 0.3, \rho = 2/6$
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	$0.3 \leq x \leq 1, \rho = 1/6$
6	$-1 \leq x < -0.3, \sigma_v = 0.05$	$-1 \leq x < -0.3, \rho = 3/6$
	$-0.3 \leq x < 0.3, \sigma_v = 0.1$	$-0.3 \leq x < 0.3, \rho = 2/6$
	$0.3 \leq x \leq 1, \sigma_v = 0.4$	$0.3 \leq x \leq 1, \rho = 1/6$
7	0.1	$-1 \leq x < -0.3, \rho = 7/10$
		$-0.3 \leq x < 0.3, \rho = 2/10$
		$0.3 \leq x \leq 1, \rho = 1/10$

Tabela 4.2: Número de dados de treinamento em cada conjunto.

Conjunto	Número de pontos
1	100
2	1000
3	100
4	1000
5	100
6	1000
7	100

Tabela 4.3: PCIP para cada conjunto de treinamento.

Conjunto	S-VInet	E-VInet	S-MLP	E-MLP
1	96.703	97.419	88.417	97.067
2	96.199	96.308	90.874	96.096
3	96.001	97.043	95.368	98.794
4	95.568	96.200	93.235	95.549
5	87.326	94.370	86.232	98.911
6	93.100	96.206	87.388	94.492
7	89.100	97.712	89.920	96.068

tornam mais sensíveis às variações de ruído. Os Conjuntos 3 e 4 são casos fáceis para todas as redes, pois apresentam mais dados nos locais onde existem mais ruído. Conjuntos 5 e 6 são casos difíceis para a S-VInet, mas a E-VInet fornece um PCIP melhor porque incorpora a incerteza das regiões de baixa densidade. É importante notar que o aumento no número de pontos de treinamento faz o PCIP da E-MLP melhorar em todos os casos. O Conjunto 7 mostra como o IP se comporta com dados com ruído constante e densidade variável no espaço de entrada, ambas S-VInet e S-MLP obtiveram resultados semelhantes, mas a E-VInet e E-MLP deram resultados melhores.

A Figura 4.1 mostra o IP para o Conjunto 1. S-VInet e E-VInet obtiveram resultados parecidos mas uma diferença visível existe entre a S-MLP e E-MLP. Enquanto a S-MLP não consegue refletir as variações de ruído, a E-MLP se comporta como a VInet e fornece um IP mais estreito próximo de -1 e mais largo próximo de +1, devido ao aumento do ruído nos dados de treinamento.

A Figura 4.2 mostra os resultados para o Conjunto 5. Esse é um teste difícil para todas as redes, porque a densidade é mais baixa onde existe mais ruído e a rede não consegue uma boa aproximação. A E-VInet calcula o IP melhor porque é mais sensível as variações de densidade, E-MLP obteve um PCIP razoável mas claramente superestimou o IP para a região próxima de +1.

A Figura 4.3 mostra o IP para o Conjunto 7, onde o ruído é constante e a densidade variável. Novamente o IP da S-MLP ficou constante e não reflete a variação da densidade, enquanto que o da E-VInet e E-MLP apresenta um IP mais condizente com a variação da densidade, embora pareça mais suave com a E-VInet.

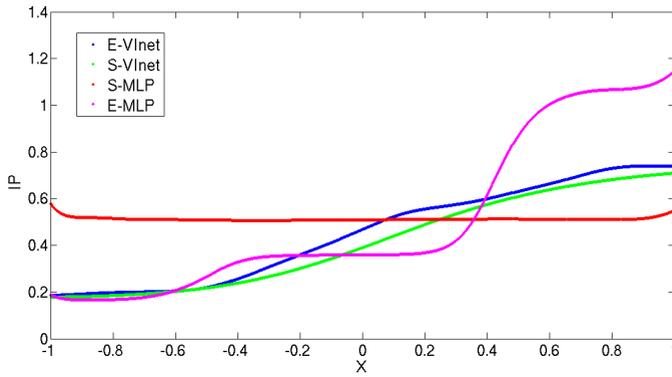


Figura 4.1: IP para o Conjunto 1, dados de treinamento com 3 regiões com ruído diferente.

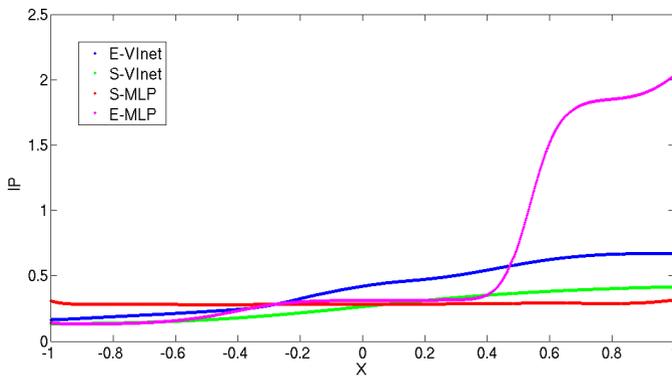


Figura 4.2: IP para o o Conjunto 5, dados de treinamento com diferentes níveis de ruído e densidade, a região com mais ruído possui menor densidade.

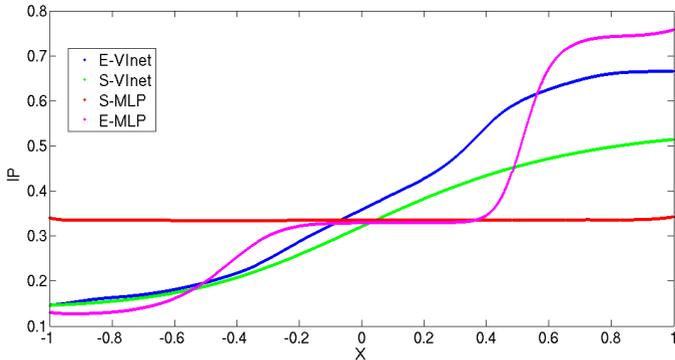


Figura 4.3: IP para o Conjunto 7, ruído constante mas regiões com densidade diferente.

4.3 EXPERIMENTO 2

Um segundo exemplo foi feito para comparar as metodologias apresentadas em um espaço de entrada multidimensional. Esse é um problema real, retirado da base de dados Proben1 (PRECHLT; INFORMATIK, 1994), e tem como característica ser um problema de múltiplas entradas, mas uma única saída de previsão de consumo de energia em uma construção e baseado em 14 variáveis de entrada. A base de dados completa tem 4208 padrões. Todas as 100 redes VInet possuem 40 neurônios na camada escondida e a rede SOM foi criada com uma grade de 10 por 10 neurônios. As MLPs foram criadas com 10 neurônios na camada escondida.

Três testes foram desenvolvidos: o primeiro com 50% dos dados foram aleatoriamente selecionados para treinamento e o restante para testes; no segundo 25% dos dados foram utilizados para treinamento e os outros 75% para testes; no terceiro foi utilizada a mesma configuração do segundo teste mas removendo-se dados especificamente de dois agrupamentos, criando assim áreas de densidade bem baixa.

Os resultados estão na Tabela 4.4. Pode-se perceber que no primeiro teste todas as abordagens apresentaram resultados semelhantes, embora a E-VInet superestimou um pouco, por incorporar no IP a informação da densidade. O segundo teste mostra que a estratégia de aproximação local empregada pela VInet apresenta resultados melhores quando se possui menos dados. O PCIP mudou pouco do primeiro para o segundo teste com as VInets, mas causou um forte impacto nas MLPs. Isso ocorre porque algumas regiões com baixa densidade foram criadas e o IP da MLP não consegue identificar a distribuição dos pontos. No

Tabela 4.4: PCIP para o segundo experimento, o primeiro teste possui 2104 pontos escolhidos aleatoriamente, o segundo teste possui 1052 pontos também escolhidos aleatoriamente e o terceiro possui 740 pontos.

	S-VInet	E-VInet	S-MLP	E-MLP
1° Teste	95.501	97.288	90.522	92.035
2° Teste	92.046	98.179	76.086	75.520
3° Teste	75.687	90.234	67.853	69.496

terceiro teste, como alguns agrupamentos foram removidos, deixando algumas regiões do espaço de entrada sem dados. A S-VInet obteve bem menos de 95% no PCIP por não incorporar a informação da densidade. As MLPs também obtiveram resultados ruins pelo mesmo motivo do segundo teste, o cálculo do IP não incorpora as incertezas em relação a baixa densidade dos dados de treinamento.

4.4 EXPERIMENTO 3

O terceiro experimento é um problema geofísico que envolve a aplicação de redes neurais para caracterização de reservatórios. Nesse tipo de aplicação, os especialistas da área têm que enfrentar a seguinte situação: existem alguns poços exploratórios perfurados em coordenadas específicas de uma região de prospecção. Através dos *logs* dos poços é possível obter os valores de atributos sísmicos e propriedades petrofísicas ao longo da perfuração. Esses registros formam o conjunto de treinamento para uma rede neural. Posteriormente, atributos sísmicos extraídos a partir de outros dados sísmicos são usados como entrada para estimar propriedades petrofísicas em torno de um reservatório. Análise de incerteza e estimação de intervalos de confiança desempenham um papel crítico na indústria de petróleo e exploração de gás e pode até mesmo determinar a viabilidade financeira da exploração de um reservatório. Um intervalo de confiança grande em alguma região avaliada pode indicar a extrapolação, dados de treinamento esparsos ou mesmo alta quantidade de ruído nos dados, indicando que mais poços devem ser perfurados naquelas coordenadas ou que os métodos de aquisição de dados devem ser revistos.

Neste trabalho utilizou-se como atributo de entrada para a rede neural a variação da impedância de onda compressional, ou seja, a variação da impedância ao longo de um determinado intervalo de tempo, chamado de *delta* impedância-P. Tem-se a intenção de correlacioná-la com a variação da saturação de água do reservatório como propriedade petrofísica desejada. Impedância-P é a propriedade física do que representa a resistência de uma rocha para a propagação de uma onda compressional (NEVES; ROISENBERG; NETO, 2009),

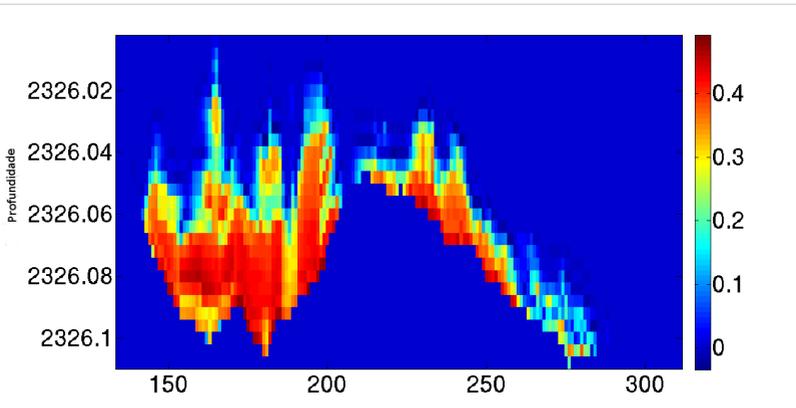


Figura 4.4: Saída desejada no problema de caracterização de reservatório.

Tabela 4.5: PCIP para o problema de caracterização de reservatórios, o conjunto de treinamento possui 748 pontos e o de teste possui 4806.

	S-VInet	E-VInet	S-MLP	E-MLP
Experimento 3	91.548	95.963	89.825	93.695

que pode ser definido como o produto entre a velocidade da onda compressional e densidade da rocha. A saturação de água é uma propriedade petrofísica que indica o volume de água contida nos poros da rocha e sua variação ao longo de um determinado período de tempo, é a propriedade do reservatório a ser caracterizada. A Figura 4.4 mostra a saída desejada desse experimento.

Neste experimento foi utilizado um conjunto de dados de 10 *logs* de poços localizados em uma região modelada na Bacia de Santos e uma seção sísmica para realizar a caracterização de uma parte do reservatório. A seção sísmica corresponde a uma fatia retirada de um volume sísmico, que contém dados sobre os atributos sísmicos e propriedades petrofísicas, organizada para fornecer uma representação espacial da região de interesse.

Um total de 100 redes de cada modelo foram treinados os dados de treinamento. As MLPs e VInets foram criados com três neurônios na camada escondida. A Figura 4.5 mostra os dados de treinamento e podemos ver três regiões com diferentes níveis de ruído.

O PCIP para a S-MLP ficou abaixo do desejado por não identificar a variação do ruído, entretanto a E-MLP calcula IPs mais largo onde o ruído e a incerteza são maiores. Como pode ser visto na Figura 4.6 o IP da S-MLP foi quase constante no intervalo entre $[-1, 1]$ onde estão os dados de treinamento e

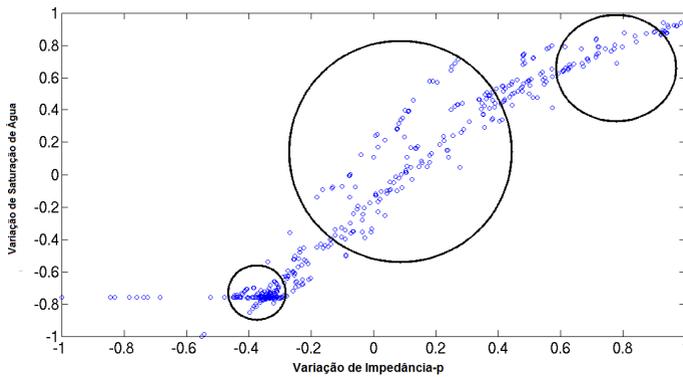


Figura 4.5: Conjunto de treinamento no problema de caracterização de reservatório. Os círculos mostram as regiões com diferentes níveis de ruído.

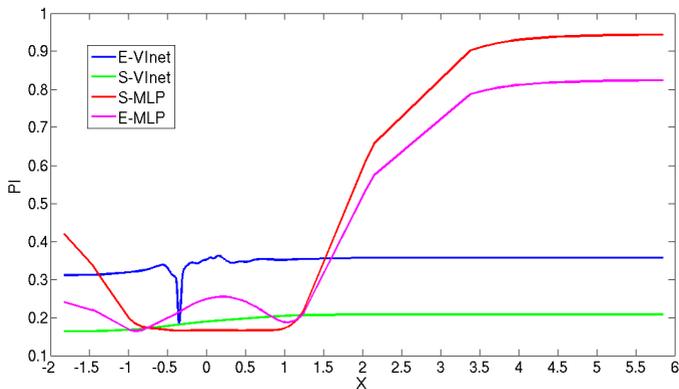


Figura 4.6: IP para o problema de caracterização de reservatórios.

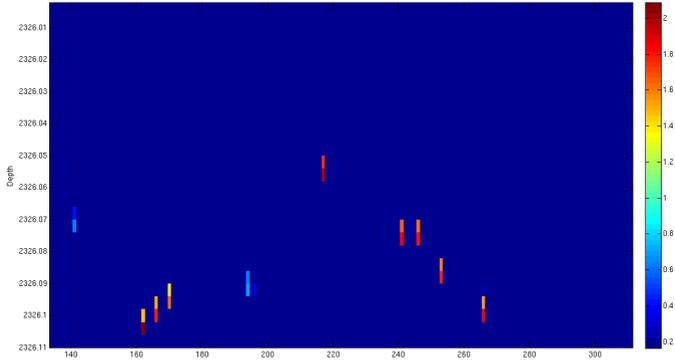


Figura 4.7: IP da E-VInet exibindo como uma seção vertical.

ficou mais largo nas regiões de extrapolação menor que -1 e maior que 1 , por outro lado a E-MLP percebeu que os dados no intervalo $[-1, 1]$ não possuem ruído constante. O IP da E-VInet mostra um comportamento curioso, a densidade dos dados de treinamento na região ao redor de -0.4 é alta, então o IP ficou estreito nessa região. VInets foram menos sensíveis ao ruído na região $[-1, 1]$ e o IP aumentou um pouco. Uma clara desvantagem da VInet é que seu IP não percebeu as regiões de extrapolação.

Figura 4.7 mostra o IP da E-VInet como uma fatia vertical do reservatório. As regiões onde o IP é maior são regiões de extrapolação na Figura 4.6. Isso é um bom indicativo sobre a incerteza da rede e pode significar que é necessário obter mais dados de treinamento nessa região para se obter mais confiança na saída da rede.

5 CONCLUSÕES

Este trabalho analisou o comportamento do intervalo de predição da VInet sob diferentes configurações de ruído e densidade, que não foram testadas no trabalho original de Leonard et al. (LEONARD; KRAMER; UNGAR, 1992), como por exemplo as diversas configurações do Experimento 1 do Capítulo de Testes e Resultados. Outros trabalhos que compararam as medidas de confiança de redes neurais também não chegaram a comparar o desempenho da rede com vários conjuntos de teste (YANG et al., 1991).

Foram propostas mudanças para corrigir as deficiências que o modelo exibiu em relação ao cálculo da densidade da VInet, que além de poder causar problemas de arredondamento em dados multidimensionais, também não é otimizado para estimar a densidade e sim para a aproximação de função. A incorporação da densidade no IP também se mostrou útil para torná-lo mais próximo do desejado e ajuda a refletir melhor a incerteza existente no conjunto de treinamento, como Yang et. al (YANG et al., 1991) notou, o IP VInet por si só não consegue capturar as variações da densidade (RODRIGUES-NETO; ROISENBERG; SCHWEDERSKY-NETO, 2009), (RODRIGUES-NETO; ROISENBERG; SCHWEDERSKY-NETO, 2010).

Por fim, as modificações posposta foram comparadas em testes de problemas reais e artificiais, tanto com a VInet como com redes MLPs, que também apresentam técnicas para estimar medidas de confiança na saída. As diferenças intrínsecas entre a MLP e a RBF, devido a abordagem de aproximação global da MLP e local na RBF resulta em características diferentes no cálculo do IP. Por exemplo, a proposta para o cálculo do IP para a MLP de Chryssolouris et. al (CHRYSSOLOURIS; LEE; RAMSEY, 1996) não consegue capturar a variação do ruído no espaço de entrada, o que pode prejudicar a utilização da rede em problemas reais, como no Experimento 3, que os dados reais do problema de caracterização de reservatórios exibem três regiões distintas de ruído.

5.1 LIMITAÇÕES E TRABALHOS FUTUROS

Um dos problemas enfrentados nesse trabalho foi como utilizar uma boa métrica para avaliar o IP, a maioria dos trabalhos relacionados avalia apenas o PCIP (PAPADOPOULOS; EDWARDS, 2001), (DONALDSON; SCHNABEL, 1987), (HWANG; DING, 1997), (SHRESTHA; SOLOMATINE, 2006), mas o formato do IP também pode ser relevante para o problema.

Outra limitação está na equação que adiciona a densidade no IP, foi estimado que nas regiões de densidade mais baixa o IP deve ser o dobro do que

é calculado normalmente, mas essa heurística pode não ser verdadeira para todos os problemas. Uma maneira de determinar automaticamente a influência da densidade no IP poderia ser interessante para resolver esse problema.

Para pesquisas futuras, seria de grande interesse usar a rede SOM para estimar o IP, além da densidade. Alguns trabalhos usam redes neurais para calcular empiricamente o IP (SHRESTHA; SOLOMATINE, 2006), (CHINNAM; BARUAH, 2007), utilizando como entrada os resíduos da fase de treinamento. Essa informação poderia ser adicionada junto à densidade estimada pela SOM.

Também é interessante a discussão sobre as causas de problemas de aprendizagem das redes neurais. Neste trabalho as causas dos problemas de aprendizagem foram abordados como um problema de engenharia e identificado como principais causas o ruído e a baixa densidade de exemplos de treinamento, porém outras áreas de conhecimento podem ter abordagens diferentes para problemas de aprendizagem.

Em relação a SOM, alguns trabalhos otimizam a rede para melhorar o casamento da densidade, como por exemplo o *algoritmo de consciência* que pune os neurônios que mais vencem e possibilita que outros neurônios que nunca venceram a competição possam vencer também (DESIENO, 1988). Existem várias outras modificações nas redes SOM que podem trazer resultados melhores para densidade (HULLE; LEUVEN, 1998), é interessante um estudo comparativo entre as vantagens e desvantagens dessas técnicas.

Bibliografia

BISHOP, C. M. Novelty detection and neural network validation. In: *IEE Proc.-Vis. Image Signal Process., Vol. 141*. [S.l.: s.n.], 1994.

BISHOP, Christopher M. *Neural Netowrks for Pattern Recognition*. [S.l.]: Oxford Press, 1995.

BROOMHEAD, D.S.; LOWE, D. Multivariable functional interpolation and adaptative networks. In: *Complex Systems*. [S.l.: s.n.], 1988. v. 2, p. 321–355.

CHEN, S. Nonlinear time series modelling and prediction using gaussian rbf networks with enhanced clustering and rls learning. *Electronics Letters*, v. 31, n. 2, p. 117–118, jan. 1995. ISSN 0013-5194.

CHEN, S.; BILLINGS, S. A.; LUO, W. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, Taylor & Francis, v. 50, p. 1873–1896, 1989. Disponível em: <<http://eprints.ecs.soton.ac.uk/1147/>>.

CHEN, S.; COWAN, C. F. N.; GRANT, P. M. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, IEEE, v. 2, n. 2, p. 302–309, 1991. Disponível em: <<http://eprints.ecs.soton.ac.uk/1135/>>.

CHEN, S.; MULGREW, B.; MCLAUGHLIN, S. Adaptive bayesian decision feedback equaliser based on a radial basis function network. *Conference record, SUPERCOMM/ICC '92, Discovering a New World of Communications., IEEE International Conference on Communications*, v. 3, p. 1267–1271 vol.3, jun. 1992.

CHINMAN, R.B.; DING, J. Prediction limit estimation for neural network models. *Neural Networks, IEEE Transactions on*, v. 9, n. 6, p. 1515–1522, Nov 1998. ISSN 1045-9227.

CHINNAM, Ratna Babu; BARUAH, Pundarikaksha. Empirical prediction limit estimation methods for feed-forward neural networks. *International Journal of General Systems*, v. 36, n. 2, p. 221–236, 2007.

CHRYSSOLOURIS, G.; LEE, M.; RAMSEY, A. Confidence interval prediction for neural network models. In: *IEEE Trans. Neural Networks* 7. [S.l.: s.n.], 1996. p. 229–232.

CICHOCKI, A.; UNBEHAUEN, R. *Neural Networks for Optimization and Signal Processing*. [S.l.]: John Wiley & Sons, 1993.

DESIENO, D. Adding a conscience to competitive learning. *Neural Networks, 1988., IEEE International Conference on*, p. 117–124 vol.1, jul. 1988.

DONALDSON, Janet R.; SCHNABEL, Robert B. Computational experience with confidence intervals for nonlinear least squares. *Technometrics*, American Society for Quality Control and American Statistical Association, Alexandria, Va, USA, v. 29, n. 1, p. 67–82, 1987. ISSN 0040-1706.

DUDA, R. O.; HART, P. E. *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, 1973.

GIROSI, F.; PODDIO, T. Networks and the best approximation property. In: *Biological Cybernetics* 63. [S.l.: s.n.], 1990. p. 169–176.

HARTMAN E. J., J. D. Keeler; KOWALSKI, J. M. Layered neural networks with gaussian hidden units as universal approximations. In: *Neural Computation* 2. [S.l.: s.n.], 1990. p. 210–215.

HAYKIN, Simon. *Neural Networks: A Comprehensive Foundation*. [S.l.]: Prentice Hall, 1998.

HESKES, Tom. Practical confidence and prediction intervals. In: *Advances in Neural Information Processing Systems* 9. [S.l.]: MIT press, 1997. p. 176–182.

HOLMSTROM, Lasse; HAMALAINEN, Ari. The self-organizing reduced kernel density estimator. In: *IEEE International Conference on Neural Networks*. [S.l.: s.n.], 1993. p. 417–421.

HULLE, M.M. Van; LEUVEN, K.U. Clustering with kernel-based equiprobabilistic topographic maps. *Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop*, p. 204–213, aug. 1998. ISSN 1089-3555.

HWANG, J. T. G.; DING, A. A. Prediction intervals for artificial neural networks. In: *J. American Statistical Association* 92(438). [S.l.: s.n.], 1997. p. 748–757.

J. ROGERS J.A., Rogers J.R. Wiener; R., Moll. Predicting carbonate permeabilities from wireline logs using a back-propagation network. In: *61st SEG meeting*. [S.l.: s.n.], 1991.

KOHONEN, Teuvo. The self-organizing map. *Proceedings of the IEEE*, v. 78, n. 1, 1990.

LEONARD, J. A.; KRAMER, M. A.; UNGAR, L. H. A neural network architecture that compute its own reliability. In: *Computers Chem. Engng.* [S.l.: s.n.], 1992.

LIPPMANN, R.P. Pattern classification using neural networks. *Communications Magazine, IEEE*, v. 27, n. 11, p. 47–50, 59–64, nov. 1989. ISSN 0163-6804.

LORENZETTI, E. Predicting lithology from vp and vs using neural networks. In: *62st SEG meeting*. [S.l.: s.n.], 1992.

LUTTRELL, Stephen P. Code vector density in topographic mappings: Scalar case. *IEEE Transactions on Neural Networks*, v. 2, 1991.

MCKAY, D. J. C. A practical bayesian framework for backprop networks. In: *Neural Computation*. [S.l.: s.n.], 1992. p. 448–472.

MOODY, J.; DARKEN, C. J. Fast learning in networks of locally tuned processing units. In: *Neural Computation 1*. [S.l.: s.n.], 1989. p. 281–294.

MUNOZ, A.; SANZ-BOBI, M. A. An incipient fault detection system based on the probabilistic radial basis function network. application to the diagnosis of the condenser of a coal power plant. In: *Neurocomputing*. [S.l.: s.n.], 1998. p. 177–194.

NEAL, Radford M. *Bayesian Learning For Neural Networks*. [S.l.]: Springer-Verlag, 1996.

NEVES, Cicero Augusto Magalhaes da Silva. Um método para construir intervalos de predição sensível ao ruído em redes neurais. *Dissertação de Mestrado na Universidade Federal de Santa Catarina*, 2009.

NEVES, Cicero Augusto Magalhaes da Silva; ROISENBERG, Mauro; NETO, Guenther Schwedersky. A method to estimate prediction intervals for artificial neural networks that is sensitive to the noise distribution in the outputs. *IEEE Press, Piscataway, NJ, USA*, p. 2608–2612, 2009.

- NIX, D. A.; WEIGEND, A. S. Learning local error bars for nonlinear regression. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 1995.
- ORR, Mark. Introduction to radial basis function networks. *Technical Reports*, v. 1, 1996.
- ORR, Mark. Combining regression trees and rbfs. *International Journal of Neural Systems*, v. 10, 1999.
- PAPADOPOULOS, Georgio; EDWARDS, Peter J. Confidence estimation methods for neural networks a practical comparison. In: *IEEE Transactions on Neural Networks*. [S.l.: s.n.], 2001.
- PARK, J.; SANDBERG, I. W. Universal approximation using radial basis function networks. In: *Neural Computation* 3. [S.l.: s.n.], 1991. p. 246–257.
- PARK, J.; SANDBERG, I. W. Approximation and radial basis function networks. In: *Neural Computation* 5. [S.l.: s.n.], 1993. p. 305–316.
- PARZEN, E. On estimation of a probability density function and mode ann. In: *Math. Statist.* 33. [S.l.: s.n.], 1962. p. 1065–1076.
- POWEL, M.J.D. Radial basis functions for multivariable interpolation: a review. Claredon Press, v. 1, p. 143–167, 1987.
- PRECHELT, Lutz; INFORMATIK, Fakultat Fur. *PROBEN1 - a set of neural network benchmark problems and benchmarking rules*. [S.l.], 1994.
- RITTER, H. Asymptotic level density for a class of vector quantization processes. *IEEE Transactions on Neural Networks*, v. 2, 1991.
- RODRIGUES-NETO, Abner Cardoso; ROISENBERG, Mauro; SCHWEDERSKY-NETO, Guenther. Análise e extensão de uma rede neural capaz de calcular sua própria confiabilidade. In: . [S.l.: s.n.], 2009.
- RODRIGUES-NETO, Abner Cardoso; ROISENBERG, Mauro; SCHWEDERSKY-NETO, Guenther. Efficient confidence bounds for rbf networks for sparse and high dimensional data. *ICANN 2010: Proceedings of the 2010 international conference on Artificial Neural Networks*, 2010.
- SHAO, R. et al. Confidence bounds for neural network representations. In: *Computers chem. Engng* 21. [S.l.]: Elsevier Science Ltd, 1997. p. 1173–1178.
- SHRESTHA, Durga; SOLOMATINE, Dimitri. Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, Elsevier, v. 19, n. 1, 2006.

SPECHT, Donald F. Probabilistic neural networks. *Neural Netw.*, Elsevier Science Ltd., Oxford, UK, UK, v. 3, n. 1, p. 109–118, 1990. ISSN 0893-6080.

TERRELL, George R.; SCOTT, David W. Variable kernel density estimation. In: *The Annals of Statistics*, Vol. 20, No. 3. [S.l.]: Institute of Mathematical Statistics, 1992. p. 1236–1265.

TRAVEN, H.G.C. A neural network approach to statistical pattern classification by ‘semiparametric’ estimation of probability density functions. *Neural Networks, IEEE Transactions on*, v. 2, n. 3, p. 366–377, may. 1991. ISSN 1045-9227.

TSAI JI-ZHENG CHU, Shi-Shang Jang Po-Feng; SHIEH, Shyan-Shu. Developing a robust model predictive control architecture through regional knowledge analysis of artificial neural networks. In: *Journal of Process Control*. [S.l.: s.n.], 2003. p. 423–435.

VEAUX, Richard De; SCHWEINSBERG, Jason; SHELLINGTON, David. Prediction intervals for neural networks via nonlinear regression. In: *Computers chem. Engng 21*. [S.l.]: Elsevier Science Ltd, 1998. p. 1173–1178.

VOUMVOULAKI, Emmanouil M.; HATZIARGYRIOU, Nikos D. A particle swarm optimization method for power system dynamic security control. In: *IEEE Transactions on Power Systems*. [S.l.: s.n.], 2010. p. 1032–1041.

WEDDING, Donald K.; CIOS, Krzysztof J. Certainty factors versus pazen windows as reliability measures in rbf networks. In: *Neurocomputing 19*. [S.l.]: Elsevier Ltd, 1997. p. 151–165.

WEDDING-II, Donald K.; CIOS, Krzysztof J. Time series forecasting by combining rbf networks, certainty factors, and box-jenkins model. In: *Neurocomputing*. [S.l.: s.n.], 1996. p. 149–168.

YANG, Luren et al. An evaluation of confidence bound estimation methods for neural networks. In: *ESIT 2000*. [S.l.: s.n.], 1991.