Aproximação de funções com intervalos de confiança

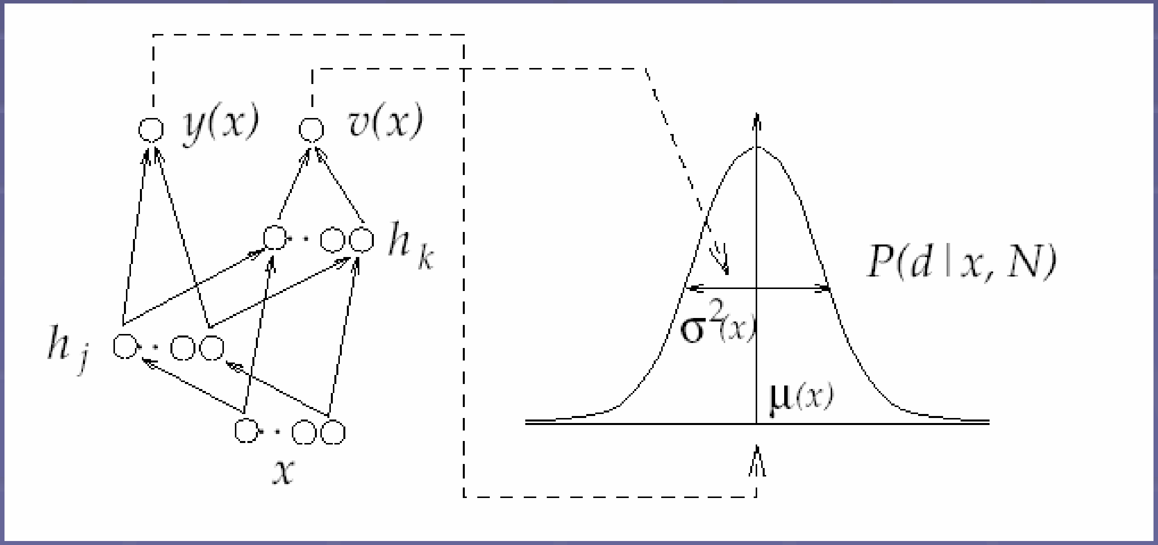
Adreas S. Weingend
David A. Nox

Resumo para a disciplina Redes Neurais.
Professor Antonio Carlos Thomé

Adriano Martins Moutinho

Obtendo barras de erro usando um esquema de máxima semelhança

- Usar redes neurais não só para estimar a média de uma função *f(x)*, mas também para estimar *v(x)*, a variância de *f(x)*.
- Utiliza-se, para isto, uma camada de neurônios auxiliar que estima a variância da função, escolhe-se uma função de ativação exponencial para este neurônio.



O problema do alvo

- O alvo para as unidades de saída é facilmente obtido, pois é a função *f(x)* que se deseja estimar.
- No entanto, o alvo para os neurônios que estimam a variância não está disponível, sendo então criado um método de atualização dos pesos destas camadas maximizando o negativo do Log da semelhança ou custo, dado por:

$$C = -\sum_{i} \ln P(x_i / d_i, \eta)$$

Modelo Gaussiano de Erro

- Assumindo um modelo gaussiano para o erro, ou seja, que f(x) será aproximada por d(x) + n(x).

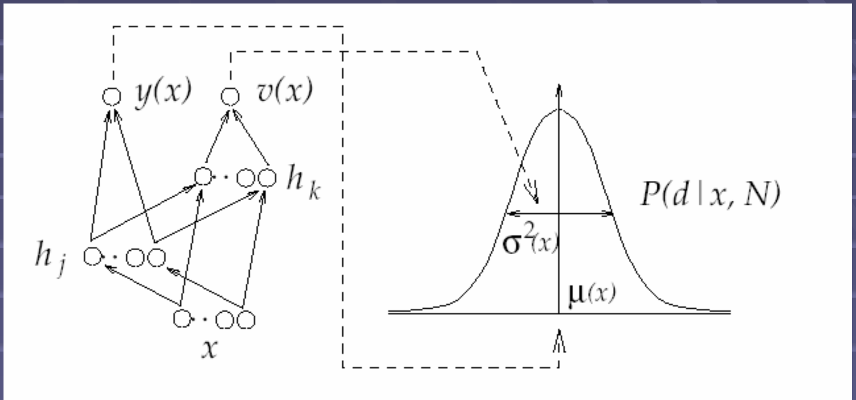
$$P(x_i/d_i, N) = \left[\frac{1}{\sqrt{2\pi\sigma^2(x_i)}}\right] \cdot e^{-\left\{\frac{[d_i - \mu(x_i)]^2}{2\sigma^2(x_i)}\right\}}$$

$$\therefore -\ln(P(x_i/d_i, N)) = \frac{2\pi v(x_i)}{2} + \left[\frac{[d_i - y(x_i)]^2}{2v(x_i)}\right]$$

$$C = -\sum \ln P(x_i/d_i, N) = \sum \frac{1}{2}\left(\frac{[d_i - y(x_i)]^2}{v(x_i)} + \ln[v(x_i)]\right)$$

Backpropagation para camada de estimativa de variância

- Para a camada *y* temos: (derivando o custo C em respeito aos pesos)



$$\Delta w_{yj} = \eta \cdot \frac{1}{v(x_i)} \left[d_i - y(x_i) \right] h_j(x_i)$$

- Para a camada v temos:

$$\Delta w_{yj} = \eta \cdot \frac{1}{2 \cdot v(x_i)} \left[\left[d_i - y(x_i) \right]^2 - v(x_i) \right] h_j(x_i)$$

Modificação feita na taxa de aprendizado, baixando em faixas onde a variância é alta!

Mecânica de treinamento

- Como a camada de estimativa da variância faz a atualização dos pesos de acordo com uma estimativa da variância dos padrões de entrada. O treinamento é dividido em três partes para evitar valores imprecisos de variância, até que *y(x)* (saída da rede) esteja pelo menos um pouco aproximado à *f(x)*. O treinamento é dividido em três partes.

Mecânica de treinamento (cont.)

- Fase 1 (Aprendendo a média):

  Os dados são divididos aleatoriamente em duas partes iguais A e B. Faz-se o treinamento usando gradiente descendente simples, ignorando os termos de variâncias na equações de atualização de pesos, e usando A como treinamento e B como validação cruzada.

Mecânica de treinamento (cont.)

- Fase 2 (Aprendendo a variância):

  Usando agora B como treinamento e A como validação cruzada, se conecta a camada *v* e treina-se usando como função de performance os erros quadráticos. Usa-se, para tal, ainda um gradiente descendente simples ignorando os termos de variância nas atualizações de pesos. Os pesos obtidos na fase 1 ficam congelados.

Mecânica de treinamento (cont.)

- Fase 3 (Regressão ponderada):

  Divide-se novamente o conjunto de dados em A’ e B’, descongela-se todos os pesos e usa-se o treinamento conforme descrito anteriormente, ou seja, usando-se os termos de variância que impedem que o ajuste dos pesos seja grande em regiões onde a variância é grande.

Experimento #1

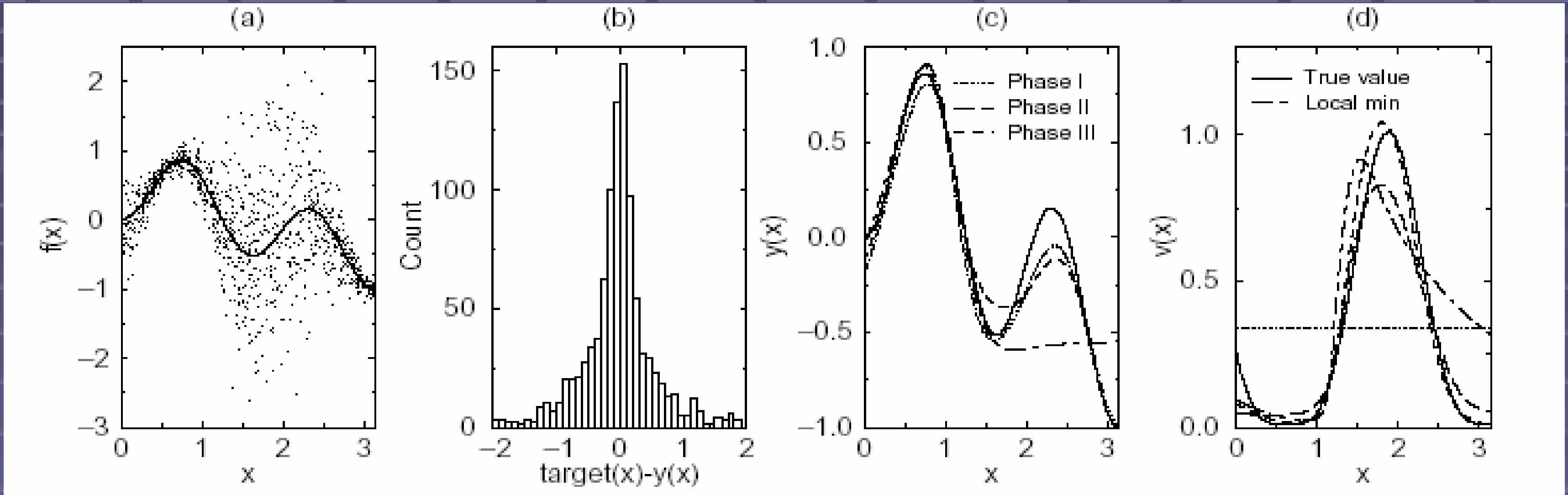
- Gera-se dados a partir de uma f(x) conhecida, tendo uma variância $\sigma^2(x)$ também conhecida.

$$f(x) = \sin(w_\alpha x) \cdot \sin(w_\beta x) \rightarrow w_\alpha = 2{,}5 \; e \; w_\beta = 1.5$$

- O alvo é gerado adicionando-se um ruído gaussiano com variância em função de x.

$$\sigma^2(x) = 0{,}01 + 0{,}25 \cdot \left[1 - \sin(w_\alpha x)\right]^2$$

Experimento #1 (cont.)



Conclusões:

- O uso das três fases permitiu uma melhor aproximação da função *f(x)* na faixa onde o ruído é menor, ignorando a faixa onde há maior incerteza. Obteve-se também, uma melhor aproximação da variância de *f(x)*.

Experimento #2

- Banco de dados da competição “Santa Fé” de intensidade de lasers.
- O objetivo da competição era estimar, com barras de erro, uma continuação de 100 pontos para a série de intensidade de lasers.
- Foram usados métodos para reamostrar a série e obter mais do que os 1000 pontos disponíveis.

Experimento #2 (cont.)

- Usando-se o método descrito no paper, foram executadas as etapas 1, 2 e 3 sem nenhuma alteração significativa.
- No entanto, para melhorar a performance da rede, foi utilizada uma técnica de validação que previa 100 pontos de continuação da série a partir de 500 pontos contínuos retirados de um local aleatório dos dados.

Experimento #2 (cont.)

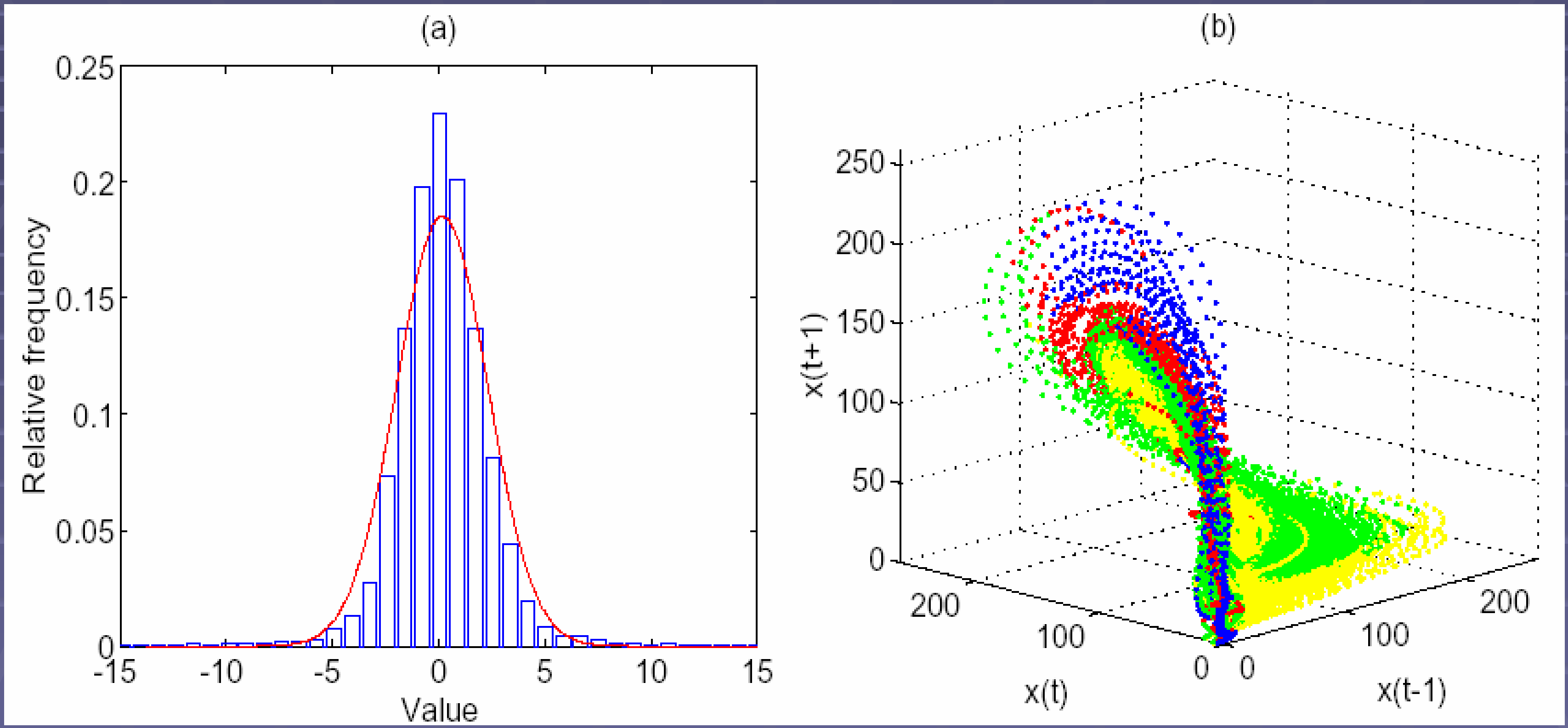
- Assim, foram geradas duas continuações para a série. Uma chamada “single step” e outra iterativa.
- Na série “single step” a continuação é gerada automaticamente pela rede.
- Na continuação iterativa, faz-se previsões para a série usando ± uma variância do vetor de entrada. Se a distância entre estas duas previsões for maior que a estimativa de variância dada pela rede, esta distância é usada como estimativa de erro.
- Também foi colocada um limite em 4 na variância prevista pois na competição era muito crítico definir uma barra de erros muito pequena.

Experimento #2 (cont.) Resultados:

	Train		Test		
	900-pt single st.	900–1000 iterated	1–100 single st.	1–100 iterated	
	NMSE	NMSE	NMSE	NMSE	NALL
Naive Baseline	—	—	—	**1.001**	**8.50**
Sauer (1994)	—	—	—	**0.080**	**4.84**
Wan (1994)					
Network	0.0004	0.0026	0.0230	**0.055**	—
Competition Entry	—	—	—	**0.028**	**3.48**
This Paper					
Network	0.0010	0.0142	0.0198	**0.096**	—
Competition Conditions	—	—	—	**0.016**	**3.28**

- Os resultados foram melhores que os dois primeiros colocados na competição

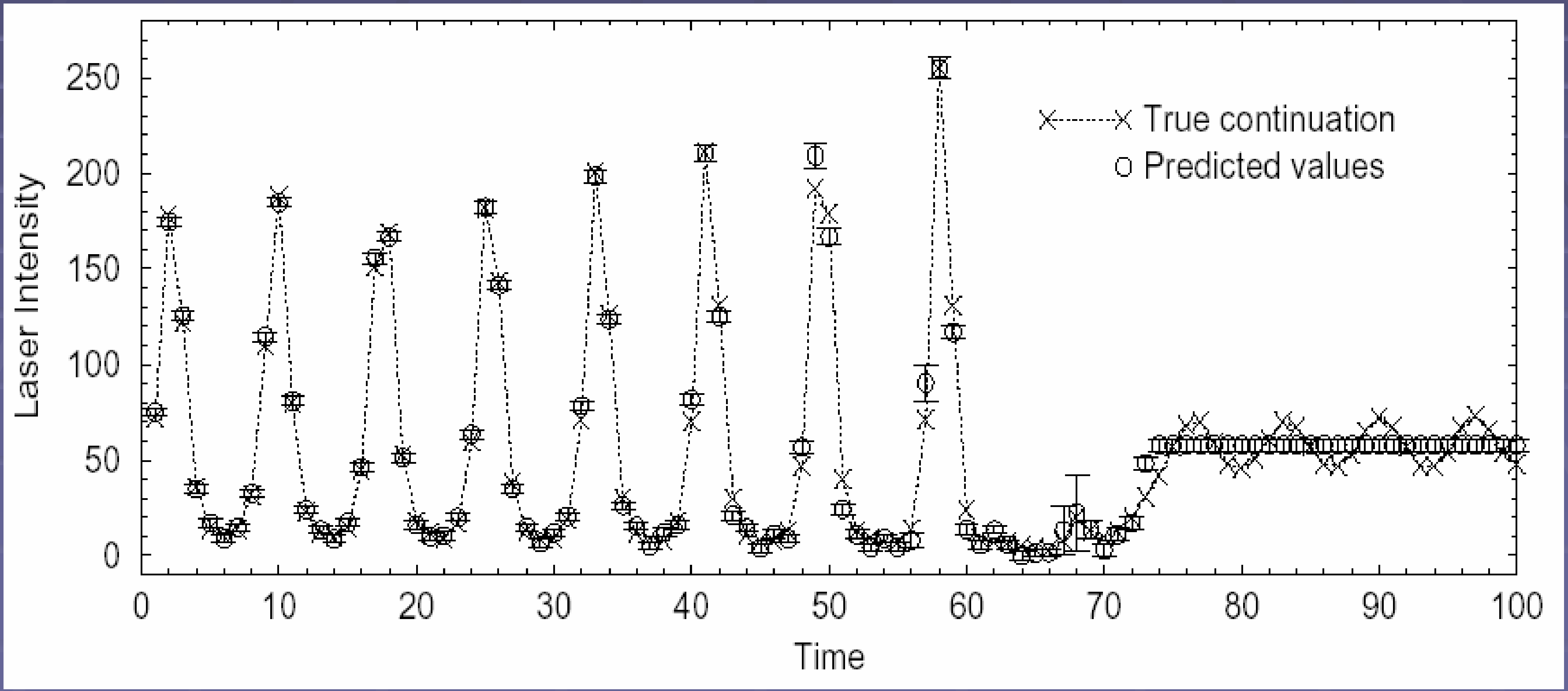
Resultados



- O histograma decai rapidamente, o que indica um bom desempenho do sistema.
  A não simetria do histograma indica variância não uniforme

Conclusões:

- A série foi bastante aproximada pela técnica até próximo a t = 66s, onde ocorre um colapso e os valores passam a não ser mais previsíveis. Neste ponto a previsão passa a ser feita com base na média e variância do ponto de treinamento onde isso também ocorre.



Conclusões finais

- Este método pode ser usado como método matemático de previsão da variância de uma série temporal, tendo-se uma previsão dos pontos onde a série não será previsível.
- Com alguns ajustes, pode-se utilizar os métodos descritos neste paper em problemas de classificação de padrões, onde se terá a classe a qual pertence o padrão, bem como uma quantificação da incerteza deste fato.