

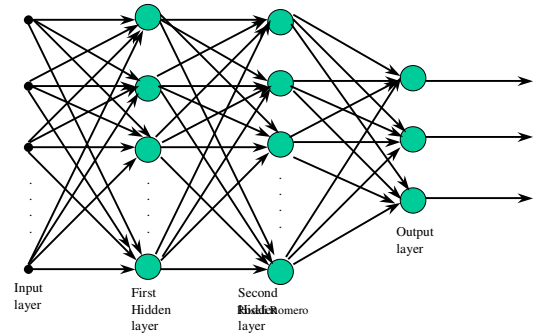
SCE 5809 – REDES NEURAIAS

REDE NEURAL DO TIPO MULTI-CAMADAS

Profa. Roseli Romero

Roseli Romero

Modelo de Rede Neural com Multiplas Camadas



II - Algoritmo Back-Propagation

$$\text{Out}(x) = g(\sum_j W_j g(\sum_i w_i x_i))$$

Isto é uma função não-linear de uma combinação linear de funções não lineares de combinações lineares das entradas

Roseli Romero

II - Algoritmo BackPropagation

OBJETIVO

- Encontrar um conjunto de pesos $\{W_j\}, \{w_{jk}\}$, para

MINIMIZAR $\sum_i (y_i - \text{Out}(x_i))^2$ pelo metodo do “gradiente descent”.

OBS: Convergência para um MINIMO global não é garantida.

Na prática: não é problema!!!

Roseli Romero

II - Algoritmo Back-Propagation

$$\Delta w_{jk} = -\eta \delta_j^p \text{out}_k^p$$

- Se o neurônio está na camada de saída

$$\delta_{pj} = (y_j^p - \text{out}_j^p) f'(net_j^p) \quad net_j^p = \sum_k w_{jk} \text{out}_k^p$$

- Se o neurônio está na camada oculta

$$\delta_{pj} = f'(net_j^p) \sum_k \delta_k^p w_{kj}$$

Roseli Romero

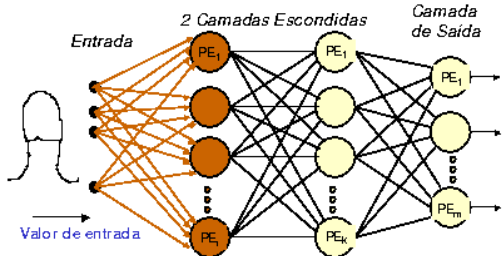
MULTI-LAYER PERCEPTRON

- Redes de apenas uma camada só representam funções linearmente separáveis
- Redes de múltiplas camadas solucionam essa restrição
- O desenvolvimento do algoritmo Back-Propagation foi um dos motivos para o ressurgimento da área de redes neurais em 1986 por Rumelhart et.

Roseli Romero

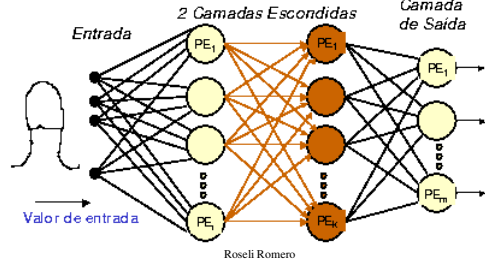
PROCESSO DE APRENDIZADO

Fase 1: *Feed-Forward* Fluxo de Dados



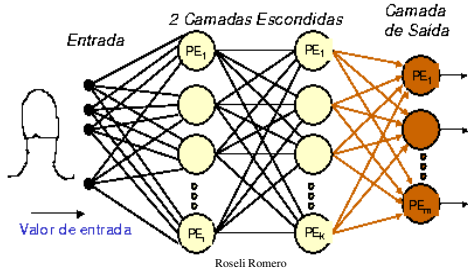
PROCESSO DE APRENDIZADO

Fase 1: *Feed-Forward* Fluxo de Dados



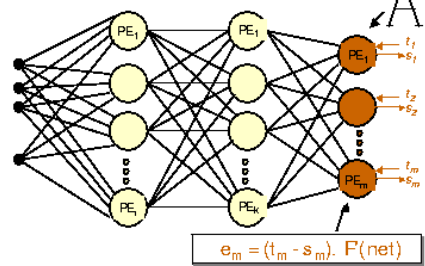
PROCESSO DE APRENDIZADO

Fase 1: *Feed-Forward* Fluxo de Dados



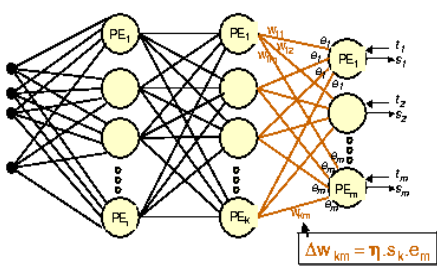
PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
Cálculo do erro da camada de saída



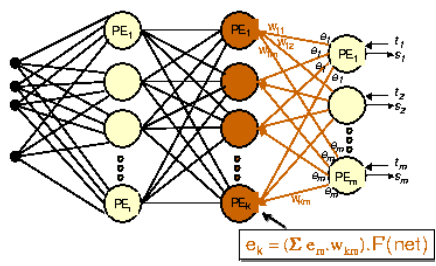
PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
Atualização dos pesos da camada de saída



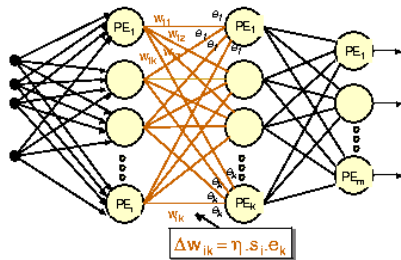
PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
Cálculo do erro da 2ª camada escondida



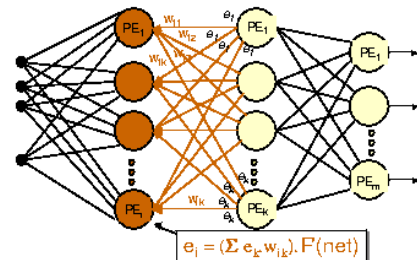
PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
 Atualização dos pesos da 2ª camada escondida



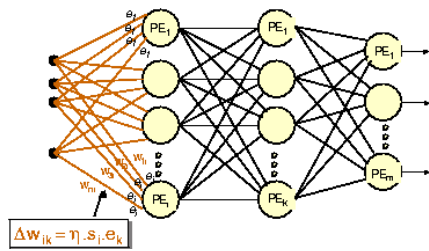
PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
 Cálculo do erro da 1ª camada escondida



PROCESSO DE APRENDIZADO

Fase 2: *Feed-Backward* Fluxo de Erros
 Atualização dos pesos da 1ª camada escondida



Roseli Romero

ALGORITMO

Este procedimento de aprendizado é repetido diversas vezes, até que *para todos processadores de camada de saída e para todos padrões de treinamento*, o erro seja menor do que o especificado.

Roseli Romero

ALGORITMO

Inicialização: pesos iniciados com valores aleatórios e pequenos ([-1,1])
 Treinamento

Repita

Considere um novo padrão de entrada x_i e seu respectivo vetor de saída t_i desejado do conj. de treinamento;

Repita

- Aplica-se o mesmo par (x_i, t_i)
- calcule se as saídas dos processadores, começando da primeira camada escondida até a camada de saída;
- calcula-se o erro na camada de saída
- atualiza os pesos de cada processador, começando pela camada de saída, até a camada de entrada;

até que erro quadrático médio para esse padrão, seja \leq tolerância.

até que o erro quadrático médio seja \leq tolerância
 para todos os padrões de conjunto de treinamento

Roseli Romero

RNA para Compressão de Imagens

- Padrão JPEG: mais utilizado
- PCA Clássica: método estatístico multivariado
- Rede PCA Adaptativa: arquitetura de Redes Neurais Artificiais

Roseli Romero

Motivação

- Redução da quantidade de dados armazenadas em sistemas computacionais
- Redução da dimensionalidade de imagens que ocupem grande quantidade de memória
- Obtenção de métodos que
 - atinjam altas taxas de compressão
 - não prejudiquem a qualidade visual

Roseli Romero

RNA PCA Adaptativa

- Postulado de Hebb[Hebb-1949]: “Quando um axônio da célula A está suficientemente próximo para excitar uma célula B e repetidamente tenta excitá-la, algum processo crescente ou mudanças metabólicas ocorrem em ambas as células”.
- Transformando em regras [Sten-73]:
 - Se dois neurônios ligados por uma sinapse são simultaneamente ativados, a intensidade dessa sinapse(conexão) é aumentada
 - Se dois neurônios ligados por uma sinapse são ativados assincronamente, a intensidade dessa sinapse é diminuída ou até mesmo eliminada.

Roseli Romero

RNA PCA Adaptativa

- Regra de Hebb:

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n)$$

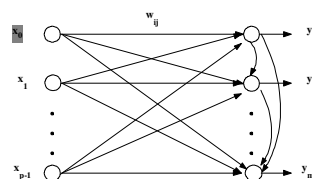
Onde η é uma constante positiva – veloc.de aprendizado
 $y_k(n)$ é a saída do neurônio k no tempo n
 x_j é j-ésimo elemento do vetor de entrada no tempo n

- Regra Anti-Hebbiana [Foldiak, 1989]

$$\Delta w_{kj}(n) = -\eta y_k(n) x_j(n)$$

Roseli Romero

Rede PCA Adaptativa



$$\text{Saída: } y_j(n) = \sum_{i=0}^{p-1} w_{ij}(n) x_i(n) + \sum_{l < j} u_{lj}(n) y_l(n),$$

$$\text{Ajuste dos pesos: } \Delta w_{ij}(n) = \eta x_i(n) y_j(n)$$

$$\text{Ajuste dos pesos laterais: } \Delta u_{lj}(n) = -\mu y_l(n) y_j(n)$$

Roseli Romero

Teorema de Convergência[Sang-89]

“Se a matriz de pesos sinápticos $W(n)$ for associada a valores aleatórios no tempo $n = 0$, então, com probabilidade 1, a regra generalizada de Hebb irá convergir na média, e, no limite, irá se aproximar de uma matriz cujas colunas serão os primeiros m autovetores da matriz C de covariância dos vetores de entrada $x(n)$, ordenados por ordem decrescente de autovalor.”

Roseli Romero

Teorema

Portanto, no limite, pode-se escrever:

$$\Delta w_j(n) \rightarrow 0 \quad W_j \rightarrow a_j \quad j = 0, 1, \dots, m-1$$

tal que $\|w_j(n)\| = 1$ para todo j . Os valores representam os autovetores normalizados associados aos m maiores autovalores da matriz C de covariância dos vetores de entrada $x(n)$, estando esses autovalores ordenados em ordem decrescente.

Roseli Romero

- Pode-se acelerar a convergência da rede introduzindo um termo *momentum* β e deixando que os parâmetros de aprendizagem e o *momentum* diminuam com o tempo. Em [MAO-95], utilizou-se:

$$\Delta w_{ij}(n+1) = \eta(n) x_i y_j + \beta(n) \Delta w_{ij}(n) \quad (1)$$

Roseli Romero

$$\Delta u_{ij}(n+1) = -\mu(n) y_i y_j + \beta(n) \Delta u_{ij}(n) \quad (2)$$

onde $\eta(n+1) = \max(\alpha\eta(n), 0.0001)$, $\mu(n+1) = \max(\alpha\mu(n), 0.0002)$, $\beta(n+1) = \max(\alpha\beta(n), 0.0001)$ e α é o fator de diminuição.

Roseli Romero

Algoritmo PCA Adaptativa-[Rubn-89]

Início

1. Inicialize todos os pesos de conexões com pequenos valores aleatórios e escolha os valores para os parâmetros de aprendizagem. Normalize-os em [0,1]. Se normalizar em [-1,1] pode mudar os sinais dos auto-vetores.

2. Repita

2.2 Selecione aleatoriamente um padrão p -dimensional e apresente-o à rede.

2.3. Ajuste os pesos das conexões entre a camada de entrada e a camada de saída de acordo com a Eq. (1)

2.4. Normalize os vetores-peso (em colunas).

2.5. Atualize os pesos laterais de acordo com a equação (2) (não precisa normalizar)

2.6. Modifique os parâmetros β , η , e μ .

Até que {todos os pesos laterais sejam suficientemente pequenos (a soma de seus valores absolutos seja menor que algum *threshold* ϵ) ou {um número de iterações máximo seja atingido}.

Fim.

Roseli Romero

Experimentos Realizados

- Dados sobre as imagens:
 - conjunto composto por 208 imagens médicas
 - dimensão 480x640 pixels (valor em nível de cinza)
 - representação de cortes de um fígado humano
 - aquisição a partir de um microscópio laser, pelo Departamento de Patologia da Fiocruz

Roseli Romero

Taxas de Compressão

$$\text{Taxa de Compressão} = 1 - \frac{\text{comprimento da cadeia de dados comprimidos}}{\text{comprimento da cadeia de dados originais}}$$

Taxas de Compressão para Blocos 32x32

Número de Componentes Principais	Taxa de Compressão
1	$1 - \left(\frac{1}{32}\right) = 96,875\%$
2	$1 - \left(\frac{2}{32}\right) = 93,75\%$
3	$1 - \left(\frac{3}{32}\right) = 90,625\%$

Roseli Romero

Compressão de Imagens através do Padrão JPEG

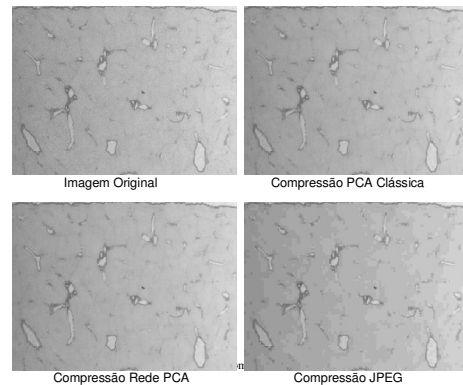
- Utilização do aplicativo xv (xview) para linux:
 - possui módulo para compressão JPEG
 - possibilita que o usuário selecione taxa de compressão

Roseli Romero

Discussão de Resultados (1)

MSE	Taxa de Compressão								
	Rede PCA Adaptativa			PCA Clássica			JPEG		
	96,875%	93,75%	90,625%	96,875%	93,75%	90,625%	96%	93%	90%
Imagem 1	5.82	4.91	4.27	5.43	4.69	4.21	9.72	7.01	6.10
Imagem 2	6.75	5.49	4.78	6.17	5.17	4.62	9.97	7.58	6.65
Imagem 3	5.20	4.08	3.40	4.64	3.77	3.26	8.91	6.44	5.39
Imagem 4	5.80	4.36	3.58	5.18	4.07	3.44	9.30	6.72	5.64
Imagem 5	6.08	4.50	3.67	5.30	4.16	3.50	9.47	6.76	5.65
Imagem 6	5.19	4.00	3.34	4.64	3.75	3.22	8.90	6.44	5.36
Imagem 7	5.41	3.96	3.33	4.81	3.74	3.21	9.04	6.51	5.39
Imagem 8	5.39	4.06	3.39	4.79	3.81	3.26	8.99	6.48	5.37
Imagem 9	5.32	3.92	3.27	4.66	3.68	3.16	8.93	6.40	5.32
Imagem 10	5.61	4.22	3.53	Roseli Romero	3.97	3.41	9.47	6.63	5.57

Discussão de Resultados (2)



Conclusões

- 3 técnicas de compressão foram apresentadas e aplicadas em uma sequência de imagens médicas
- Analisando os MSE cometidos pelas técnicas,
 - Os resultados da Rede PCA foram bem similares aos obtidos pela PCA Clássica
 - O desempenho do padrão JPEG foi inferior aos desempenhos obtidos pelas duas outras técnicas

Roseli Romero

Referências

- Haykin, S., "Neural Networks", IEEE Press, 1999
- RUMELHART, D.; MCCLELLAND and THE PDP RESEARCH GROUP - Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. I. Cambridge MA:MIT Press, 1986.
- BRAGA, Antônio de Pádua.; Carvalho, André Carlos Ponce de Leon Ferreira. "Redes neurais artificiais : teoria e aplicações", Rio de Janeiro : LTC Editora, 2000. 262 p.

Roseli Romero