

1 - Absorção do escalamento:

1 - Na rede neural abaixo os neurônios 1 e 3 são lineares, e o neurônio 2 é tipo $tgh(\cdot)$. As variáveis de entrada e saída tem distribuição aproximadamente Gaussiana, com parâmetros conforme a tabela 1 abaixo. Após o treinamento com as entradas e saídas normalizadas para média zero e desvio padrão unitário as sinapses obtidas para a rede estão apresentadas na tabela 2 abaixo. Apresente os valores das sinapses a serem usadas com a rede com as entradas originais, não normalizadas.

Obs: Examine com cuidado o caso das sinapses conectadas ao neurônio 1, que ao mesmo tempo pertence à camada intermediária e fornece uma saída. Aplique a denormalização para as entradas x , em seguida para a saída y_a e finalmente como neurônio da camada intermediária para a denormalização de y_b .

Tabela 1

	μ	σ
x_a	5	2
x_b	10	3
y_a	4	2
y_b	10	3

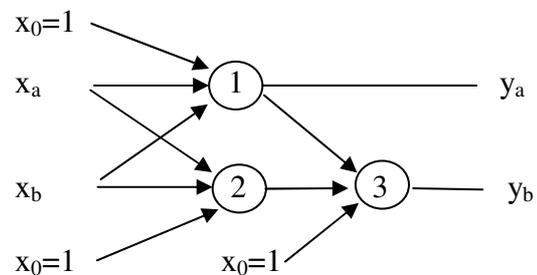


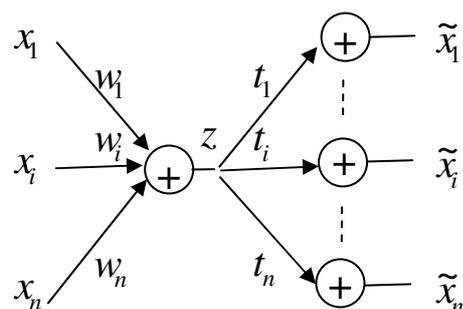
Tabela 2 - Sinapses W_{ij}

$i \downarrow j \Rightarrow$	a	b	0	1	2
1	.5	1.1	.7	-	-
2	-.3	.2	0	-	-
3	-	-	.2	.3	-.3

2 - PCA generalizada

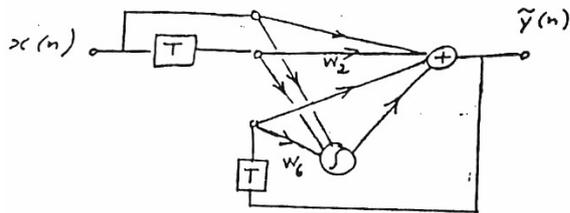
2.1 - Um mapeamento linear $\underline{y} = \underline{M} \underline{x}$ deve passar por uma compressão $\underline{z} = \underline{W} \underline{x}$ e pela reconstrução $\tilde{\underline{y}} = \underline{T} \underline{z}$ tal que minimize o erro $F = E \left\{ \sum_{\forall l} a_l (y_l - \tilde{y}_l)^2 \right\}$. Mostre como fazê-lo usando redes neurais. Discuta a relação entre os mapeamentos \underline{W} e \underline{T} .

2.2 - Para o exercício acima, considerando que as variáveis x_i estão normalizadas para média zero e desvio padrão unitário, apresente as sinapses que deverão ser utilizadas para entradas e saídas não normalizadas.



3 - Modelagem

3.1 - A figura abaixo mostra a estrutura da rede utilizada para simular um sistema dinâmico não-linear, quando em operação. A rede utiliza neurônios sem bias, um linear (+) e um tipo tgh (J) e atrasos unitários T.

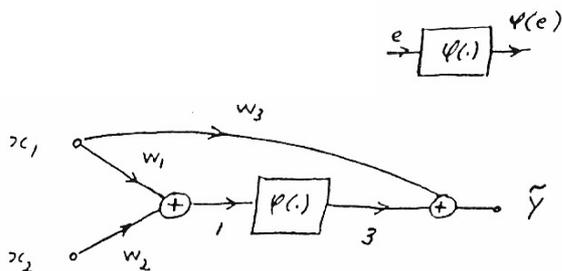


Parte da tabela usada no treinamento (série-paralelo, regra delta com $\alpha = 0,1$) é apresentada a seguir.

n	x(n)	y(n)
⋮	⋮	⋮
48	.40	.35
49	.45	.50
50	.50	.60
51	.55	.65
52	.60	.75
⋮	⋮	⋮

Em um instante do treinamento, por acaso todas as sinapses são iguais a 0,25. É sorteado então $n = 50$, e o passo de treinamento é dado. Quais os novos valores das sinapses w_2 e w_6 ? Indique como realizou os cálculos!

3.2 - O modelo de uma planta apresentado na figura abaixo inclui um bloco cuja função de transferência $\varphi(\cdot)$ já é conhecida analiticamente e não necessita ser treinado.



deduza as fórmulas de treinamento de w_1 , w_2 e w_3 por backpropagation regra delta para minimizar o erro médio quadrático da saída y sobre um conjunto de pares de treinamento $\{ \underline{x}, y \}$.

Existem restrições sobre $\varphi(\cdot)$? Caso sim, quais?

4 - Redes RBF

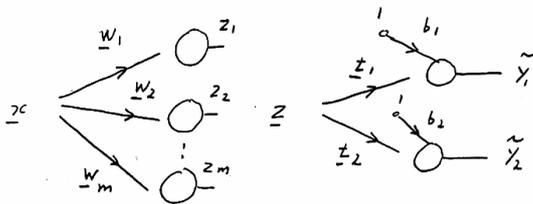
4.1 – Na rede abaixo os neurônios da camada intermediária são do tipo RBF com função de transferência

$$z_i = \frac{1}{1 + u_i} \quad \text{onde} \quad u_i = (\underline{x} - \underline{w}_i)^t \underline{K} (\underline{x} - \underline{w}_i)$$

onde \underline{K} é uma matriz diagonal com elementos k_i . Os neurônios de saída são do tipo perceptron:

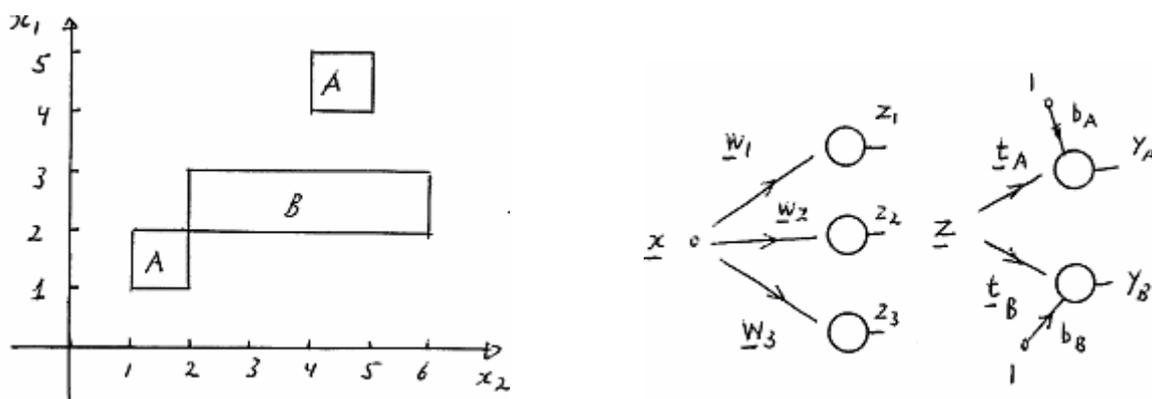
$$\tilde{y}_1 = u_1 \quad \text{onde} \quad u_1 = \underline{z}^t \underline{t}_1 + b_1 \quad e$$

$$\tilde{y}_2 = \text{tgh}(u_2) \quad \text{onde} \quad u_2 = \underline{z}^t \underline{t}_2 + b_2$$



A função objetivo a ser minimizada no treinamento é o erro médio quadrático na saída. Estabeleça as fórmulas para um treinamento tipo gradiente descendente (backpropagation).

4.2 – As duas classes abaixo devem ser separadas entre si e do restante do espaço de entrada pela rede RBF ao lado, da forma mais eficaz possível. Projete a rede.



Os neurônios da primeira camada são do tipo base radial:

$$u_i = (\underline{x} - \underline{w}_i)^t \underline{K}_i (\underline{x} - \underline{w}_i)$$

$$z_i = e^{-u_i}$$

E os da segunda camada do tipo tgh:

$$u_i = \mathbf{z}_i^t \mathbf{t}_i + b_i \quad y_i = \text{tgh}(u_i)$$

4.3 - Uma rede MLP com uma única camada com neurônios do tipo

$$u = \sum_{i=0}^{n+1} w_i x_i \quad \text{onde} \quad x_0 = 1 \quad e \quad x_{n+1} = \sum_{i=1}^n w_{n+1} x_i^2$$

$$v = \text{tgh}(u)$$

opera como um classificador com separadores esféricos.

a - Discuta a localização dos separadores em função do valor das sinapses

b - Apresente as fórmulas para o treinamento backpropagation.

c - Sugira o que usar como valores iniciais para as sinapses.

5 - Poda

5.1 - Considere uma rede feedforward de uma camada com neurônios lineares. Mostre que se a matriz de autocorrelação das entradas $\underline{R} = E(\underline{x}^t \underline{x})$ é semi-definida positiva (isto é, $\underline{w}^t \underline{R} \underline{w} \geq 0 \forall \underline{w}$ e a igualdade é atingida para pelo menos um $\underline{w} = \underline{w}_c$) é possível eliminar uma sinapse arbitrária por neurônio sem degradar o erro médio quadrático. Mostre como a partir de \underline{w}_0 (vetor sinapse ótimo de um neurônio) e \underline{w}_c encontrar o novo minimante \underline{w}_1 (novo vetor sinapse ótimo) que tem a i -ésima componente nula.

5.2 – Usando um processo oposto ao de poda é possível melhorar a robustez da rede distribuindo o processamento entre as sinapses de forma que o maior número delas tenha contribuição mais ou menos equivalente para a saída. Para isto procura-se aumentar os valores das sinapses menores e reduzir os valores das maiores usando uma função objetivo modificada

$$F = F_0 + F_1 = F_0 - \gamma \sum_{\forall i,j} w_{ij}^2 \exp\left(-\frac{w_{ij}^2}{a^2}\right)$$

onde F_0 é o erro médio quadrático na saída da rede.

a – Discuta se esta função F_1 é uma escolha adequada para o objetivo.

b – Calcule o Δw_{ij} a ser aplicado em cada sinapse.

c – Crie critérios para a escolha de γ e a .