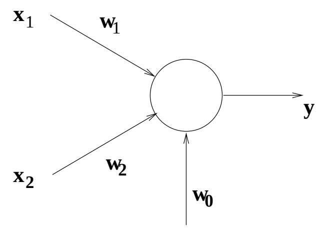
Problema 1

Considere a *adaline* apresentada na figura, com função de activação linear e sem entrada de polarização.



Admita que as entradas e saídas desejadas resultam de observações $(\mathbf{x}, d)$ de variáveis aleatórias (V.A.). Admita ainda que $\mathbf{x} = [x_1, x_2]^T$ e que as V.A. x_1 e x_2 são independentes, tendo ambas média nula e variância 2. Considere um funcional de custo definido por

$$C(\mathbf{w}) = \frac{1}{2}E[(y-d)^2]$$

e que num determinado problema de aprendizagem automática a solução óptima para o vector de pesos é dada por

$$\mathbf{w} = [w_1, w_2]^T = [1, 2]^T$$

i) Determine a correlação cruzada entre as entradas e as saídas desejadas, $E[\mathbf{x}d]$.

ii) Considere que se pretende treinar este sistema por um método de gradiente exacto. Indique qual o gradiente teórico do funcional de custo $C(\mathbf{w})$ nos pontos $\mathbf{w} = [0, 0]^T$ e $\mathbf{w} = [4, 0]^T$. Represente gráficamente no plano (w_1, w_2) a solução óptima, os dois pontos referidos e os simétricos dos vectores gradientes.

iii) Indique qual o valor do passo de adaptação óptima que deve ser utilizado em cada um dos pontos considerados para atingir o mínimo o mais rapidamente possível. Comente e, caso possivel, generalize o resultado a outros pontos do plano.

Problema 2

Admita que tem um conjunto de observações perencentes a duas classes, C_0 e C_1. Pretende-se desenvolver um sistema de classificação baseado numa rede neuronal, usando como saída desejada 0 para a classe C_0 e 1 para a classe C_1. Atendendo a que a classe C_1 corresponde apenas a 5% dos padrões, utiliza-se um conjunto de treino balanceado, com 50% de padrões de cada classe, para tentar melhorar a qualidade do classificador na região menos representada. Admita que a função sigmoidal de saída da rede é definida pela função logística

$$f(s) = \frac{1}{1 + e^{-s}}$$

e que a função de custo adoptada é o erro quadrático.

i) Em problemas de classificação, a saída de uma rede neuronal pode ser um estimativa da probabilidade *a-posteriori* da classe. Indique, justificando, se isso se verifica neste caso.

ii) Em caso negativo, derive a expressão da correcção que é necessário realizar na saída da rede para estimar correctamente aquela probabilidade, Problema 3

Considere uma adaline sem entrada de polarização cuja saída é definida por

$$y = \mathbf{w}^T \mathbf{x}$$

onde $\mathbf{w}$ é o vector de pesos. Admite que se dispõe de um dado conjunto de treino de estatísticas conhecidas Admita que $\mathbf{x} = [x_1, x_2]^T$ tem média nula e que são conhecidas as seguintes estatísticas:

$$\begin{aligned} E[x_1^2] &= 1 \\ E[x_2^2] &= 2 \\ E[x_1 x_2] &= 0.5 \\ E[x_1 d] &= 2 \\ E[x_2 d] &= 4.5 \end{aligned} \tag{1}$$

a) Determine a solução óptima para o vector de pesos que minimiza o erro quadrático médio.

b) Admita que pretende determinar o conjunto de pesos óptimos por meio de um método de gradiante exacto. Determine o vector gradiente no ponto $\mathbf{w} = [3, 3]'$.

c) Em superfícies quadráticas como a que aqui se considera, é possível determinar o mínimo do funcional de custo numa única iteração utilizando o *método de Newton*, dado por

$$\Delta \mathbf{w} = -\mathbf{H}^{-1}(\mathbf{w})\mathbf{g}(\mathbf{w})$$

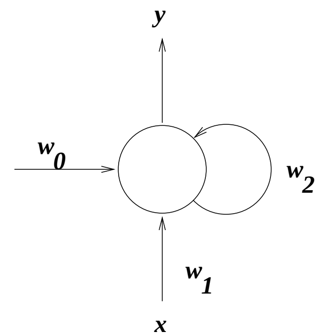
onde $H(\mathbf{w})$ é a matriz hessiana do funcional de custo no ponto considerado e $\mathbf{g}(\mathbf{w})$ o gradiente.

i) Prove que $\mathbf{H} = 2\mathbf{R}_{xx}$.

ii) Prove que, qualquer que seja o ponto inicial w_0, o método de Newton determina o mínimo do funcional de custo numa única iteração.

Problema 4

Considere a rede neuronal recorrente apresentada na figura.



Admita que a função sigmoidal da unidade da rede é definida pela função logística

$$f(s) = \frac{1}{1+e^{-s}}$$

Considere que os pesos iniciais são $w_1 = w_2 = w_3 = 1$.

i) Determine a saída da rede para a sequência de entrada $x(t)$, de dimensão 2, com $x(1) = 1$ e $x(2) = 0$.

ii) Determine os novos pesos da rede, após uma época do algoritmo de retropropagação no temp, admitindo que para a sequência de entrada $x(1) = 1$ e $x(2) = 0$, se pretende ter como sequência desejada, à saída, $d(1) = 0$ e $d(2) = 1$ e que adopta como passo de adaptação $\eta = 0.01$.

Problema 5

Admita que Admita que tem um conjunto X de padrões (vectores de dimensão n) cuja densidade de probabilidade $p(\mathbf{x})$ pretende estimar.

Indique como poderia utilizar uma rede neuronal para estimar esta densidade de probabilidade.