Redes Neuronais e Aprendizagem Automática

Exame de 5/2/04

Resolução do Problema 4

Enunciado:

Considere uma rede neuronal com uma única saída $o \in]0,1[$. Esta rede é utilizada num problema de classificação com dados pertencentes a duas classes, *A* e *B*. Durante o treino, aos dados da classe *A* corresponde o valor desejado $d = 1$, e aos da classe *B* o valor desejado $d = 0$.

Como sabe, se a função de custo que é optimizada for o erro quadrático, a saída da rede, após o treino, é um estimador da probabilidade a posteriori da classe *A*, isto é, de $P(A \mid X)$, em que *X* é o padrão de entrada. Prove que se, em vez do erro quadrático, se usar a função de custo

$$f(d,o) = \begin{cases} -\log(1-o) & \text{se } d = 0 \\ -\log o & \text{se } d = 1 \end{cases}$$

a saída da rede, após o treino, é igualmente um estimador de $P(A \mid X)$.

Notas:

- Trate apenas a situação de uma rede com um único padrão de entrada. Nesta situação, uma vez que só existe um valor de *X*, tem-se $P(A \mid X) = P(A)$. Basta, portanto, provar que nesta situação a saída é um estimador de $P(A)$.

  A extensão deste resultado para mais de um padrão de entrada poderia ser depois tratada de forma semelhante à que foi usada, nas aulas teóricas, para o erro quadrático, mas não é pedida neste exame.

- Pode admitir que $0 < P(A) < 1$, isto é, que esta probabilidade não toma os valores extremos 0 ou 1.

Resolução:

Considere-se uma rede neuronal à qual é apresentado sempre o mesmo padrão de entrada. Dado que o padrão de entrada é sempre o mesmo, a saída da rede é fixa. Designe-se essa saída por *o*. Admita-se que o padrão apresentado à rede pertence à classe *A* (correspondendo a $d = 1$) com uma probabilidade $P(A)$, e à classe *B* (correspondendo a $d = 0$) com uma probabilidade $P(B) = 1 - P(A)$. Para simplificar a notação faça-se $P(A) = p$.

O valor esperado da função de custo é dado por

$$\begin{aligned} E[f(d,o)] &= P(d=1)\,f(1,o) + P(d=0)\,f(0,o) \\ &= p\,f(1,o) + (1-p)\,f(0,o) \\ &= -p\log o - (1-p)\log(1-o) \end{aligned}$$

Numa situação ideal de treino (conjunto de treino com distribuição de probabilidades idêntica à do universo dos dados) é minimizado este valor esperado. O valor óptimo da saída, correspondente ao mínimo, pode ser obtido igualando a zero a derivada:

$$\frac{\partial E[f(d,o)]}{\partial o} = -\frac{p}{o} + \frac{(1-p)}{(1-o)}$$

Igualando a zero conclui-se facilmente que $o = p = P(A)$.

Numa situação ideal ter-se-á pois, após o treino, $o = P(A)$. Numa situação real, em que os dados de treino resultem de uma amostragem aleatória do universo, a distribuição de probabilidades no conjunto de treino aproxima a do universo mas não é, em geral, igual a ela. Numa tal situação a saída da rede será portanto um estimador de $P(A)$.