P1 – Tópicos em Aprendizagem de Máquina e Classificação de Padrões – 24/09/2010

1-Considerar a seguinte regra de decisão para um problema de classificação unidimensional com duas classes: decidir ω_1 se $x > \theta$; caso contrário decidir ω_2.

(a) mostrar que a probabilidade de erro para esta regra é

$$P(error) = P(\omega_1) \int_{-\infty}^{\theta} p(x \mid \omega_1) dx + P(\omega_2) \int_{\theta}^{\infty} p(x \mid \omega_2) dx$$

(b) por diferenciação, mostrar que a condição necessária para minimizar P(error) é que θ satisfaça:

$$p(\theta \mid \omega_1) P(\omega_1) = p(\theta \mid \omega_2) P(\omega_2)$$

(c) esta solução define θ unicamente?

2-Em problemas de classificação com várias categorias existe a opção de ou atribuir um padrão à uma das c classes, ou rejeitá-lo como sendo não reconhecível. Se o custo da rejeição não é muito alto, a rejeição pode ser uma ação aceitável. Seja

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0 & i = j \qquad\qquad i, j = 1, \ldots, c \\ \lambda_r & i = c + 1 \\ \lambda_s & \text{caso contrário} \end{cases}$$

onde λ_r é o custo de escolher a ação (c + 1), a rejeição, e λ_s o custo associado às ações restantes com $i \neq j$.

(a) mostrar que o risco mínimo é obtido pela regra de decisão:

decidir ω_i se $P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x})$ para todo j e $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$, rejeitar caso contrário.

(b) o que acontece quando $\lambda_r = 0$? E se $\lambda_r > \lambda_s$?

3-Considerar a fronteira de decisão de Bayes para um problema de classificação d dimensional com duas categorias. Mostrar que para qualquer hiperquadrática em d dimensões, existem distribuições normais $p(\mathbf{x}|\omega_i) \sim N(\mu_i, \Sigma_i)$ e probabilidades *a priori* $P(\omega_i)$, i =1,2, que possuem estas hiperquadráticas como suas fronteiras de decisão.

4-Supor que x tem uma densidade uniforme $p(x \mid \theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \textit{caso contrário} \end{cases}$

(a) dadas n amostras $\mathcal{D} = \{x_1,....,x_n\}$ obtidas independentemente a partir de $p(x|\theta)$, mostrar que a estimativa de máxima verosimilhança de θ é max $[\mathcal{D}]$, isto é, o elemento máximo de $\mathcal{D}$.

(b) supor que $n = 5$ e que o valor máximo das amostras obtidas com esta distribuição seja $\max_k x_k = 0.6$. Esboçar $p(\mathcal{D}|\theta)$ para $0 \le \theta \le 1$. Explicar porque não é necessário conhecer os valores dos outros quatro pontos.

5-O propósito deste problema é o de derivar o classificador Bayesiano para o caso d-dimensional de Bernoulli. Como de costume, considerar cada classe separadamente, interpretando $P(\mathbf{x}|\mathcal{D})$ como $P(\mathbf{x}|\mathcal{D}_i,\omega_i)$. Considerar a probabilidade condicional para uma dada categoria como sendo

$$P(\mathbf{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$$

e seja $\mathcal{D} = \{x_1,....,x_n\}$ um conjnto de n amostras obtidas independentemente a partir desta distribuição de probabilidade.

(a) se $\mathbf{s} = (s_1,...,s_d)$ é a soma de n amostras, mostrar que

$$P(\mathcal{D} \mid \boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$$

(b) assumindo uma distribuição a priori uniforme para $\boldsymbol{\theta}$ e usando a identidade

$$\int_0^1 \theta^m (1-\theta)^n \, d\theta = \frac{m!n!}{(m+n+1)!}$$

mostrar que

$$P(\boldsymbol{\theta} \mid \mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$$