Tópicos em Aprendizagem de Máquina e Classificação de Padrões – 2010

Lista de Exercícios 1: Teoria Bayesiana de Decisão

1-No caso de duas classes, o erro condicional quando se aplica a regra de decisão de Bayes é dada pela Eq. 7. Mesmo que as densidades *a posteriori* sejam contínuas, esta forma do erro condicional quase sempre leva a um integrando que não é contínuo quando se calcula o erro total conforme a Eq. 5.

(a) mostrar que para densidades arbitrárias, podemos substituir a Eq. 7 por $P(error|x) = 2P(\omega_1|x)P(\omega_2|x)$ na integral para obter um limitante superior para o erro global.

b) mostrar que se usamos $P(error|x) = \alpha P(\omega_1|x)P(\omega_2|x)$ para $\alpha < 2$, então não há garantia que a integral fornece um limitante superior para o erro global.

2-Considerar duas densidades de probabilidade na forma $p(x|\omega_i) \propto e^{-|x-a_i|/b_i}$, $i = 1, 2$, $b_i > 0$.

(a) obter a expressão analítica de cada densidade, isto é, normalizar cada função para valores arbitrários de a_i e b_i (equivalentemente, para que área no intervalo $[-\infty, \infty]$ seja igual a unidade).

(b) determinar a razão de verosimilhança em função das quatro variáveis.

(c) esboçar o gráfico da razão $P(x|\omega_1)/P(x|\omega_2)$ para $a_1 = 0$, $b_1 = 1$, $a_2 = 1$, $b_2 = 2$.

3-Um problema de classificação que tem como objetivo minimizar o valor máximo da função perda é um problema do tipo minimax. Para uma função perda binária, isto é, $\lambda_{11} = \lambda_{11} = 0$ e $\lambda_{12} = \lambda_{21} = 1$:

(a) mostrar que neste caso as regiões de classificação satisfaz

$$\int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x} = \int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x}$$

(b) esta solução é única?

8-Considerar um problema de classificação unidimensional com duas categorias cuja distribuição é Cauchy:

$$p(x \mid \omega_i) = \frac{1}{\pi b}\left(\frac{1}{1+((x-a_i)/b)^2}\right), \quad i = 1{,}2$$

(a) através de integração, verificar que as distribuições são normalizadas.

(b) supondo $P(\omega_1) = P(\omega_2)$, mostrar que $P(\omega_1|x) = P(\omega_2|x)$ se $x = (a_1 + a_2)/2$, isto é, a fronteira de decisão de erro mínimo de classificação é o ponto médio entre os picos das duas distribuições, independentemente de b.

(c) esboçar $P(\omega_1|x)$ para $a_1 = 3$, $a_2 = 5$ e $b = 1$.

(d) investigar como $P(\omega_1|x) = P(\omega_2|x)$ se comportam quando $x \rightarrow -\infty$ e $x \rightarrow \infty$.

18-Considerar duas distribuições normais unidimensionais $N(\mu_1,\sigma_1)$ e $N(\mu_2,\sigma_2)$. Sejam duas amostras aleatórias x_1 e x_2 obtidas de cada uma das distribuições normais e $x_3 = x_1 + x_2$. Suponha que se calcule esta soma repetidamente.

(a) qual é a distribuição de x_3 ?

(b) qual é a média μ_3 desta nova distribuição?

(c) qual é a variância σ_3 ?

21-Três distribuições, Gaussiana, uniforme e triangular tem, cada uma delas, média zero e variância σ^2 . Calcular e comparar as respectivas entropias.

23-Considerar uma distribuição normal tridimensional $p(\mathbf{x}|\omega) \sim N(\boldsymbol{\mu}, \Sigma)$ com

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}.$$

(a) determinar o valor da densidade no ponto $\mathbf{x}_o = (0.5, 0, 1)^t$.

(b) construir uma transformação linear de acordo com a Eq. 44. Determinar as matrizes dos autovetores e autovalores $\boldsymbol{\Phi}$ e $\boldsymbol{\Lambda}$, respectivamente. A seguir, converter a distribuição original em uma centrada na origem e com matriz de covariância unitária, $p(\mathbf{x}|\omega) \sim N(\mathbf{0}, \mathbf{I})$.

(c) aplicar a mesma transformação a $\mathbf{x}_o$ e obter o ponto $\mathbf{x}_w$.

(d) calcular a distância de Mahalanobis entre $\mathbf{x}_o$ e a média $\boldsymbol{\mu}$ para a distribuição original e entre $\mathbf{x}_w$ e a origem $\mathbf{0}$ para distribuição transformada; comparar os respectivos valores.

(e) $p(\mathbf{x}_o|N(\boldsymbol{\mu}, \Sigma)) = p(\mathbf{T}^t\mathbf{x}_o|N(\mathbf{T}^t\boldsymbol{\mu}, \mathbf{T}^t\Sigma\mathbf{T}))$ para toda transformação linear $\mathbf{T}$?

27-Supor que temos duas distribuições normais com a mesma covariância, mas média distintas, isto é, $N(\boldsymbol{\mu}_1,\Sigma)$ e $N(\boldsymbol{\mu}_2,\Sigma)$. Explicitar, em termos das probabilidades *a priori* $P(\omega_1)$ e $P(\omega_2)$, as condições para que a fronteira de decisão de Bayes (Bayes *decision boundary*) não passe entre as duas médias.

43-As componentes do vetor $\mathbf{x} = (x_1, x_2,...., x_d)^t$ são binárias, isto é, $x_i \in \{0,1\}$, $i = 1,.., d$. Seja $P(\omega_j)$ a probabilidade *a priori* do estado ω_j e $j = 1,.., c$. Definir

$p_{ij} = \Pr[x_i = 1]$ $i = 1,...., d$ $j = 1,..., c$

onde as componentes x_i são estatisticamente independentes para todo $\mathbf{x}$ de ω_j.

(a) interpretar com palavras o significado de p_{ij}.

(b) mostrar que a probabilidade de erro mínimo é atingida com a seguinte regra de decisão: decidir ω_k se $g_k(\mathbf{x}) \geq g_j(\mathbf{x})$ para todo j e k onde

$$g_j(\mathbf{x}) = \sum_{i=1}^{d} x_i \ln\frac{p_{ij}}{1-p_{ij}} + \sum_{i=1}^{d} \ln(1-p_{ij}) + P(\omega_j)$$

50-Utilizar as matrizes de probabilidade condicional do exemplo de rede Bayesiana dado em aula para responder as seguintes questões.

(a) suponha que é 20 de dezembro, fim do outono e começo de inverno, isto é, $P(a_1) = P(a_2) = 0.5$. Além disso, sabe-se que o peixe foi capturado no Atlântico Norte, isto é, $P(b_1) = 1$.Supor que a luminosidade não foi medida, mas sabe-se que o peixe é fino, isto é, $P(d_2) = 1$. Classificar o peixe como ou salmão ou robalo. Qual é a taxa esperada de erro?

(b) Supor que tudo o que se sabe é que o peixe é fino e tem luminosidade média. Qual estação é agora a mais provável? Qual é a probabilidade de se estar correto?