JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740

Homework for Module 3

1. Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution
$$p(x|\omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_i}{b}\right)^2}, \quad i = 1, 2.$$
   (a) If $P(\omega_1) = P(\omega_2)$, show that $P(\omega_1|x) = P(\omega_2|x)$ if $x = (1/2)(a_1 + a_2)$. Sketch $P(\omega_1|x)$ for the case $a_1 = 3, a_2 = 2, b = 5$. How does $P(\omega_1|x)$ behave as $x \to -\infty$? as $x \to \infty$?
   (b) Using the conditional densities in part a, and assuming equal *a priori* probabilities, show that the minimum probability of error is given by
$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_2 - a_1}{2b} \right|.$$
   Sketch this as a function of $|(a_2 - a_1)/b|$.
2. The Poisson distribution for discrete k, $k = 0, 1, 2, \ldots$ and real parameter λ is
$$P(k|\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}.$$
   (a) Find the mean of k.
   (b) Find the variance of k.
   (c) Find the mode of k.
   (d) Assume two categories C_1 and C_2, equally probable *a priori*, distributed with Poisson distributions and $\lambda_1 > \lambda_2$. What is the Bayes classification decision?
   (e) What is the Bayes error rate?
3. Let $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \sigma^2 I)$ for a two-category k-dimensional problem with $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.
   (a) Find P_e, the minimum probability of error.
   (b) Let $\mu_{\mathbf{1}} = \mathbf{0}$ and $\mu_{\mathbf{2}} = (m_1, \ldots, m_k)^T \neq \mathbf{0}$. Show that $P_e \to 0$ as the dimension k approaches infinity. Assume that $\sum_{k=1}^{\infty} m_k^2 \to \infty$.
4. Under the assumption that $\lambda_{21} > \lambda_{11}$ and $\lambda_{12} > \lambda_{22}$, show that the general minimum risk discriminant function for a classifier with independent binary features is given by $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$. What are $\mathbf{w}$ and w_0?