STAT 881
Spring 2008

HOMEWORK 1

1. Consider a multiclass problem where $y \in \{1, \ldots, k\}$ and k is the number of classes. Then a classification rule f is defined as a map from $\mathcal{X}$ to $\mathcal{Y} = \{1, \ldots, k\}$.

   (a) For the 0-1 loss, $L(f(x), y) = I(f(x) \neq y)$, show that the Bayes decision function in the multiclass problem is $f^*(x) = \arg\max_{j \in \mathcal{Y}} \eta_j(x)$, where $\eta_j(x) = P(Y = j | X = x)$.

   (b) Verify that for any classification rule $f : \mathcal{X} \to \mathcal{Y}$,
$$R(f) - R^* = E\{\max_{j \in \mathcal{Y}} \eta_j(X) - \eta_{f(X)}(X)\}.$$

   (c) Given estimates of class conditional probabilities $\hat{\eta}_j$, we have the plug-in decision function $\hat{f}(x) = \arg\max_{j \in \mathcal{Y}} \hat{\eta}_j(x)$. Prove that
$$R(\hat{f}) - R^* \leq 2E \max_{j \in \mathcal{Y}} |\eta_j(X) - \hat{\eta}_j(X)|.$$

   (d) When the misclassification costs are different, we change the loss function to $L(f(x), y) = C_{y, f(x)}$, where $C_{j,\ell}$ indicates the cost of misclassifying an observation from class j as class ℓ for $j, \ell \in \mathcal{Y}$. Show that the optimal classification rule minimizing the expected misclassification cost is
$$f(x) = \arg\min_{j \in \mathcal{Y}} \sum_{\ell=1}^{k} C_{\ell,j} \eta_\ell(x).$$

2. Let X be a d-dimensional random vector. Assume that the conditional distribution of X given $Y = 0$ is $N(\mu_0, \Sigma)$ and that of X given $Y = 1$ is $N(\mu_1, \Sigma)$. For simplicity, let $P(Y = 0) = P(Y = 1) = 1/2$. In this case, prove that the Bayes error is $\Phi(-\delta/2)$, where $\delta = \{(\mu_1 - \mu_0)^\top \Sigma^{-1} (\mu_1 - \mu_0)\}^{1/2}$ is the so-called Mahalanobis distance between the two conditional distributions, and Φ denotes the c.d.f. of the standard normal distribution.

3. Suppose that a d-dimensional random vector, $X = (X_1, \ldots, X_d)^\top$, is normally distributed with mean 0 and identity variance-covariance matrix. The class label Y is determined by the following probit model
$$Y = I(X^\top \beta + \epsilon > 0),$$
where $\epsilon \sim N(0, \sigma^2)$ and independent of X, and $\beta \in R^d$ is a fixed coefficient vector. Show that the Bayes error rate under the probit model is given by
$$R^* = \frac{1}{2} - \frac{1}{\pi} arc\ tan \sqrt{\frac{var(X^\top \beta)}{\sigma^2}}.$$
Briefly comment on the effect of σ^2 on the Bayes error.