STOR 565 Homework

1. Describe the difference between a fixed classification rule and a data based classification rule. Define and discuss the risk of rules of both types of classification rules.

2. Describe and discuss linear discriminant analysis.

3. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$ and let $Y = AX + b$ where $A \in \mathbb{R}^{l \times k}$ and $b \in \mathbb{R}^l$.

   a. Find $\mathbb{E}Y$ and $\text{Var}(Y)$.

   b. Argue carefully that Y is multinormal.

   c. Fix $v \in \mathbb{R}^l$. Using the results above, find the distribution of $U = \langle v, Y \rangle$.

4. Let X and Y be two jointly distributed random variables. Suppose that we wish to predict the value of Y based on the value of X via a function $g(\cdot)$. Suppose that we judge the quality of the prediction $g(X)$ by the expected squared error $\mathbb{E}(Y - g(X))^2$. Then it turns out that the best estimate of Y given X is the conditional expectation $E(Y \,|\, X)$.

   a. Let $g(\cdot)$ be a function. Show that $\mathbb{E}[(Y - \mathbb{E}[Y|X])\,(\mathbb{E}[Y|X] - g(X))\,|\,X] = 0$. (Hint: expand the product and use basic properties of conditional expectations.)

   b. Note that $(Y - g(X))^2 = (Y - E(Y \,|\, X) + E(Y \,|\, X) - g(X))^2$. Show by expanding the square and using the result from (a) that

$$\mathbb{E}[(Y - g(X))^2|X] \;=\; \mathbb{E}[(Y - E[Y|X])^2\,|\,X] \;+\; \mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2\,|\,X]$$

   c. Deduce from (b) that $\mathbb{E}[(Y - g(X))^2] \geq \mathbb{E}[(Y - \mathbb{E}[Y \,|\, X])^2]$

5. Consider a classification problem in which the conditional probability $\mathbb{P}(Y = 1 \,|\, X = x)$ is defined implicitly via the equation

$$\text{logit}(\eta(x : \beta)) \;=\; \beta^t \, x \tag{1}$$

where $\text{logit}(u) = \log[u/(1-u)]$ for $0 < u < 1$ is the logistic (or logit) function.

   a. Sketch the logistic function.

b. Show that, by inverting the relation (1) we have

$$\eta(x:\beta) \;=\; \frac{e^{\beta^t x}}{1+e^{\beta^t x}} \;=\; \frac{1}{1+e^{-\beta^t x}}$$

c. Consider the case that β and x are one-dimensional, and therefore real valued. Find the partial derivatives $\partial \log(\eta(x:\beta))/\partial\beta$ and $\partial^2 \log(\eta(x:\beta))/\partial^2\beta$, and show that the second partial is always negative.

6. Let $\mathcal{P} = \{f_\theta : \theta > 0\}$ be the family of exponential pdfs $f_\theta(x) = \theta e^{-\theta x}$ for $x \geq 0$. Suppose that we draw n samples independently from a fixed distribution $f_{\theta_0} \in \mathcal{P}$ and observe values $x_1, \ldots, x_n \in \mathbb{R}$.

a. Write down the likelihood $L(\theta)$ and the log-likelihood $\ell(\theta)$ for the family $\mathcal{P}$.

b. Find the maximum likelihood estimate $\hat{\theta}_n^{\mathrm{MLE}}$ of θ_0.

7. Recall from class that the Bayes rule can be written in the form $\phi^*(x) = \mathbb{I}(\delta_1(x) \geq \delta_0(x))$ where $\delta_k(x) = \log \pi_k\, f_k(x)$.

a. Show that if $f_k = \mathcal{N}(\mu_k, \Sigma_k)$ then

$$\delta_k(x) = -\frac{1}{2}x^t\Sigma_k^{-1}x + \langle x, \Sigma_k^{-1}\mu_k\rangle - \frac{1}{2}\left[\log(2\pi)^d\pi_k^{-2} + \det(\Sigma_k) + \mu_k^t\Sigma_k^{-1}\mu_k\right]$$

b. Show that if $\Sigma_0 = \Sigma_1 = \Sigma$ then the decision boundary of the Bayes rule is a hyperplane

$$B = \{x : x^t\,\Sigma^{-1}(\mu_1 - \mu_0) + (c_0 - c_1) = 0\}$$

where c_0, c_1 are constants that do not depend on x.

c. Show that if $\Sigma_0 \neq \Sigma_1$ then the decision boundary of the Bayes rule is a quadratic surface of the form

$$B = \left\{x : -\frac{1}{2}x^t(\Sigma_1^{-1} - \Sigma_0^{-1})x + x^t(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0) + (c_0 - c_1) = 0\right\}$$

8. Let X be a standard normal random variable and let $Y = X^2$.

a. Using the cdf method, find the density of Y.

b. Are X and Y independent? Why or why not?

c. What is $\mathrm{Cov}(X, Y)$? What do these results reveal about the relationship between covariance and independence?