Homework I: Stat 246

Due, Thursday, April 10

1. Consider the following familiy of decision rules for a two category problem with $x \in R$, with conditional densities denoted $f(x|Y=1), f(x|Y=2)$. For any threshold τ, decide $\hat{y}=1$ if $x > \tau$; otherwise decide $\hat{y}=2$. Show that

$$P(error) = P(Y=1)\int_{-\infty}^{\tau} f(x|Y=1)dx + P(Y=2)\int_{\tau}^{\infty} f(x|Y=2)dx.$$

Show that a *necessary* condition to minimize $P(error)$ is that τ satisfies:

$$f(\tau|Y=1)p(Y=1) = f(\tau|Y=2)p(Y=2).$$

Does this define τ uniquely? Give an example where a value of τ satisfying this criterion maximizes the probability of error.

2. If a and b are non-negative, show that $\min(a,b) \le \sqrt{ab}$. Use this to show that the error rate for a two category Bayes classifier must satisfy

$$P(error) \le \sqrt{P(Y=1)P(Y=2)} \int \sqrt{f(x|Y=1)f(x|Y=2)}dx \le \frac{1}{2}\int \sqrt{f(x|Y=1)f(x|Y=2)}dx.$$

3. Let $f(x|Y=i) \sim N(\mu_i, \sigma^2), i=1,2$ be a two category one-dimensional problem with $P(Y=1) = P(Y=2) = 1/2$. Show that the error of the Bayes classifier is

$$P(error) = \frac{1}{\sqrt{2\pi}}\int_a^{\infty} e^{-u^2/2}du,$$

where

$$a = \frac{|\mu_1 - \mu_2|}{2\sigma}.$$

Show that this goes to zero as a goes to infinity.

4. Let $X = (X_1, X_2)$ be bivariate normal with mean zero and covariance matrix

$$C = \begin{pmatrix} v_1 & a \\ a & v_2 \end{pmatrix}.$$

In class we provided a general formula for the conditional distribution $f(x_2|x_1)$. Here is an alternative derivation:

(a) Write $X_2 = \alpha X_1 + Z$ where $cov(Z, X_1) = 0$. Provide an explicit formula for α and Z.

(b) Explain why Z is Gaussian and independent of X. Write the variance of Z.

(c) Since Z is independent of X, conditional on $X_1 = x$ we have $X_2 = \alpha x + Z$. Deduce the conditional mean and variance of X_2 given $X_1 = x$.

5. Let $X_1 \sim N(\mu, \sigma_1^2)$ and assume that conditional on X_1, $X_2 \sim N(X_1, \sigma_2^2)$.

   (a) Using moment generating functions show that X_1, X_2 are jointly normal. What are the mean and covariance matrix of this bivariate normal.

   (b) Alternative derivation. Let $Z \sim N(0, \sigma_2^2)$ be independent of X_1. Define $X_2' = X_1 + Z$. Show that (X_1, X_2') has the same distribution as (X_1, X_2). Rededuce the mean and covariance matrix of X_1, X_2.

   (c) What is the marginal distribution of X_2.

6. Let $X = (X_1, X_2)$ be bivariate normal as in 4 . Let $X_i = (X_{i1}, X_{i2}), i = 1, \ldots, n$ be an i.i.d sample from the same distribution. Let

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} X_{i1} X_{i2},$$

   be an estimate for the covariance a. Clearly $E(\hat{a}) = a$ so this is an unbiased estimate. Compute the variance of $\hat{a}$. Hint: to get $E[(X_1 X_2)^2]$ use the moment generating function.

7. Let $X \sim N(0, \theta\Sigma)$ with Σ known and only the scale θ unknown. Derive the MLE for θ.

8. Experiments in estimating the covariance of a multivariate normal.

   (a) Generate a $d \times d$ random matrix of i.i.d $N(0, 1)$ variables A (make sure it is non-singular) and compute $\Sigma = A^t A$ (which is p.d.). Let $\lambda_{max}, \lambda_{min}$, be the largest and smallest eigenvalue of Σ.

   (b) Set a vector $\mathbf{e}$ of size $d^2 \times 100$, and the vectors $\mathbf{M}, \mathbf{m}$ of size 100

   Set $d = 10$ and repeat the following for $t = 1 : 100$ (100 samples),

   - Generate a sample $X_i, i = 1, \ldots, n$ of size $n = 150$ of random vectors distributed according to $N(0, \Sigma)$.
   - Compute the MLE estimate of the covariance matrix $\hat{\Sigma} = \frac{1}{n} \sum_{i=1} X_i X_i^t$.
   - For each entry Σ_{ij} compute the standardized error

$$e_{ij} = \frac{\Sigma_{ij} - \hat{\Sigma}_{ij}}{SE_{ij}},$$

     where SE_{ij} is given by problem 6. Store the e_{ij}'s in the next d^2 entries of $\mathbf{e}$.

Compute the eigenvalues of $\hat{\Sigma}$. Store the largest and smallest eigenvalues in the t'th entry of $\mathbf{M}$ and $\mathbf{m}$ respectively.

Plot the histogram of $\mathbf{e}$ (the standardized errors of the covariance estimates).

Plot the histogram of $(\mathbf{M} - \lambda_{\mathbf{max}})/\lambda_{\mathbf{max}}$, $(\mathbf{m} - \lambda_{\mathbf{min}})/\lambda_{\mathbf{min}}$ (i.e. the relative errors of the eigenvalues). What do you conclude from these comparisons?

Repeat for $d = 10, n = 1000$.

Repeat for $d = 100, n = 150$ and $d = 100, n = 1000$.

What happens when $d > n$?