ECEN 649 Pattern Recognition – Spring 2018

Computer Project 1

Due on: Mar 22 (electronic submission)

Consider the following Gaussian models, where the dimension is d, the prior class probabilities are $P(Y=0) = P(Y=1) = \frac{1}{2}$, the means are at the points $\boldsymbol{\mu}_0 = (0, \ldots, 0)$ and $\boldsymbol{\mu}_1 = (1, \ldots, 1)$.

Model M1: Independent variables, $\Sigma_0 = \Sigma_1 = \sigma^2 I_d$.

Model M2: Positively correlated variables in blocks of 2,

$$\Sigma_0 \;=\; \Sigma_1 \;=\; \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 & 0 \\ \rho & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \rho & \cdots & 0 & 0 \\ 0 & 0 & \rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & \rho \\ 0 & 0 & 0 & 0 & \cdots & \rho & 1 \end{pmatrix}_{d\times d}$$

where d is assumed to be a multiple of 2, and $0 \leq \rho \leq 1$.

Model M3: Positively correlated variables in blocks of 3,

$$\Sigma_0 \;=\; \Sigma_1 \;=\; \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & 0 & 0 & 0 \\ \rho & 1 & \rho & \cdots & 0 & 0 & 0 \\ \rho & \rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \rho & \rho \\ 0 & 0 & 0 & \cdots & \rho & 1 & \rho \\ 0 & 0 & 0 & \cdots & \rho & \rho & 1 \end{pmatrix}_{d\times d}$$

where d is assumed to be a multiple of 3, and $0 \leq \rho \leq 1$.

Note that these are homoskedastic Gaussian models (so that the optimal decision boundary is a hyperplane in each case), but with increasingly correlated features.

Problem 1: Show that the Bayes errors for each of these models are given by:

$$\varepsilon^*_{M1} \;=\; \Phi\left(-\frac{\sqrt{d}}{2\sigma}\right), \quad \varepsilon^*_{M2} \;=\; \Phi\left(-\frac{\sqrt{d}}{2\sigma}\frac{1}{\sqrt{1+\rho}}\right), \quad \text{and } \varepsilon^*_{M3} \;=\; \Phi\left(-\frac{\sqrt{d}}{2\sigma}\frac{1}{\sqrt{1+2\rho}}\right),$$

where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-x^2/2}dx$ is the CDF of a standard normal random variable. Argue that $\varepsilon^*_{M1} < \varepsilon^*_{M2} < \varepsilon^*_{M3}$, and that the difference becomes larger as ρ increases. Give a justification of this. For $d = 6$, and $\rho = 0.2, 0.8$, plot on the same axes the optimal errors for the three models as a function of the standard deviation σ. How do you interpret the results in terms of the correlation between the features? Hint: Use the expression for the Bayes error in the homoskedastic Gaussian case, given in class.

Problem 2: For $d = 2$,$\rho = 0.2$, $\sigma = 1$, draw a sample of size $n = 10$ from models $M1$ and $M2$. Obtain the NMC, LDA, and DLDA decision boundaries corresponding to these data. Plot the data (using O's for class 0 and X's for class 1) in each case, with the superimposed decision boundaries for the NMC, LDA, DLDA classifiers. Compute the error of each classifier using the formula given in class.

Problem 3: Extend the previous problem by drawing 100 samples of sizes $n = 10, 15, 20, 25, 30, 35, 40$, $d = 6$, $\rho = 0.2$, $\sigma = 1$ for M1, M2, and M3. There is no need to plot the decision boundaries. Compute the error for the NMC, LDA, and DLDA classifiers in each case using the analytical formula. Average over the 100 samples to obtain an aproximation of $E[\varepsilon_n]$ and plot this as a function of n in each case. Interpret what you see.