ECE 492-45 Homework 6

Material Covered: Logistic Regression, Linear/Quadratic Discriminant Analysis, Type I/II Error, ROC

Useful information Formal definition for covariance matrix: The *variance-covariance matrix*, or the *covariance matrix* for simplicity, can be regarded as a generalization from the variance of a random variable X to a variation measure for a random vector $\underset{\sim}{X} = (X_1, \cdots, X_n)$. Its definition is as follows:

$$\mathrm{Cov}(\underset{\sim}{X}) = \mathbb{E}\left[(\underset{\sim}{X} - \mathbb{E}[\underset{\sim}{X}])(\underset{\sim}{X} - \mathbb{E}[\underset{\sim}{X}])^T\right]. \tag{1}$$

Note that the covariance matrix is always a square matrix resulted from a vector outer product. If we explicitly write out the covariance matrix for a length-2 random vector $\underset{\sim}{x} = (X_1, X_2)$, its elementwise expression is as follows:

$$\mathrm{Cov}(\underset{\sim}{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{bmatrix}. \tag{2}$$

For more details about its definition, refer to Section 6.3.1 of Leon-Garcia's book. For its application in the multivariate Gaussian distribution, refer to Section 6.4. You may also try to read the "covariance matrix" entry on Wikipedia, but the key information may be buried in details.

Problem 1 (20 points) [Stock Market Direction Prediction] Complete *ISLR-4.6.1–5, 4.7.10.* Be super concise when reporting the results for 4.6.1–5, and be selective when reporting the results for 4.7.10.

Problem 2 (20 points) [`Boston` Dataset Prediction] Complete *ISLR-4.7.13.*

Problem 3 (20 points) [Decision Making and Associated Probability] Complete *ISLR-4.7.6–7.*

Problem 4 (20 points) [Quadratic Discriminant Analysis (QDA)]

(a) Random variables X_1 and X_2 are two predictors/features that will be used later to generate a dataset for classification. They are related by a first-order autoregressive model defined as follows:

$$X_2 = \rho X_1 + e, \tag{3}$$

where $\rho \in (0,1)$ is a fixed constant, both X_1 and X_2 are of mean μ and variance σ^2, and $e \sim \mathcal{N}\left(0, (1-\rho^2)\sigma^2\right)$ is independent of X_1. Prove that $\mu = 0$ and $\operatorname{cov}(X_1, X_2) = \rho\sigma^2$. Justify each entry of the variance-covariance matrix for random vector (X_1, X_2) of the following form:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{4}$$

(b) Using Eq. (3), simulate 500 data points for class 1 and class 2 with following multivariate Gaussian distributions respectively:

$$(X_1^{(1)}, X_2^{(1)}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \tag{5}$$

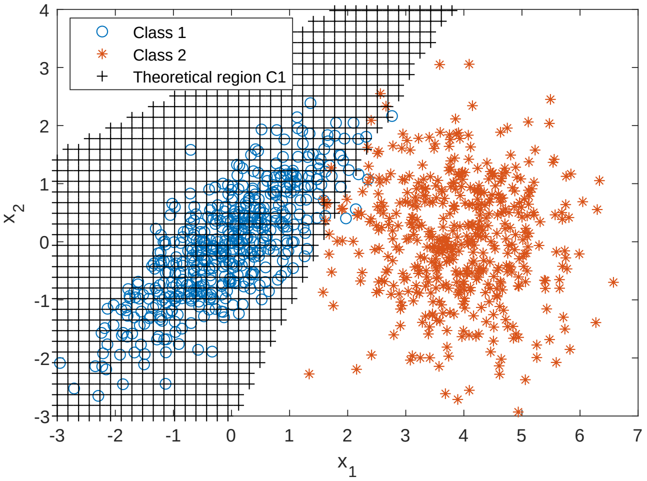
$$(X_1^{(2)}, X_2^{(2)}) \sim \mathcal{N}\left(\begin{bmatrix} d \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \tag{6}$$

where $\sigma^2 = 1$, $\rho = 0.8$, and $d = 4$. Plot all data points in plane. Use "o" for class 1 and "*" for class 2. Note that for class 2, the center is $(d, 0)$.

(c) The decision boundary is a set of 2nd-order curves shown as follows:

$$\rho c(x_1^2 + x_2^2) - 2cx_1x_2 + 2dx_1 - d^2 + \sigma^2 \ln(1-\rho^2) = 0, \tag{7}$$

where $c = \rho/(1-\rho^2)$. Use a brute-force method to draw on the data plot the theoretical decision region for class 1. Your result will be similar to the following example.



(d) (Bonus, 5 points) Start from the density function of the multivarite Gaussian, prove the expression for decision boundary given in (c). Assume equal prior.

Problem 5 (20 points) [Two-Class Discrimination]

(a) Complete *ISLR-4.7.3* to obtain a discriminant score/function $\delta_k(x)$ for class k. (Hint: After taking $\ln[\pi_k f_k(x)]$, remove all additive terms that do not contain x AND k.)

(b) Prove that the decision threshold between class 1 and class 2 is one of the roots of the following quadratic equation:

$$\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)x^2-2\left(\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}\right)x+\left(\frac{\mu_1^2}{\sigma_1^2}-\frac{\mu_2^2}{\sigma_2^2}\right)+2\ln\left(\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)=0. \tag{8}$$

(c) (Attempt this after Monday's lecture) Let $\mu_1=0$, $\mu_2=5$, $\sigma_1=1$, $\sigma_2=2$, and assume equal prior for both classes. Simulate 5,000 points for each class. Draw histograms for the two classes in the same plot. What is the theoretical decision threshold according to (b)? Use this decision threshold to calculate a confusion matrix. Clearly indicate the dimensions for ground truth and for predicted results, respectively. What are the False Positive Rate (FPR) and False Negative Rate (FNR)?

(d) (Attempt this after Monday's lecture) Generate an ROC curve by varying threshold from the small value to the largest value of the overall dataset. (Hint: Your ROC curve should tell you that at 10% false positive, the true positive rate should be from 96%–98%.)