Introduction to Statistical Learning Homework 2

Due: 2021.03.16

Lecturer: Prof. Sheng Yu

1 Problem 1

Start from the statistical decision theory

$$\hat{y}(x) = \arg\max_k P(Y = k \mid x),$$

derive the linear discriminant function

$$\delta_k(x) = \log \pi_k + x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k,$$

in LDA, and the classification rule is

$$\hat{y}(x) = \arg\max_k \delta_k(x).$$

2 Problem 2

Suppose we have features $x \in \mathbb{R}^p$, a two-class response, with class sizes N_1, N_2, and (instead of a 0/1 coding) the target coded as $-N/N_1$, N/N_2.

2.1

Show that the LDA rule classifies to Class 2 if

$$x^T \Sigma^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2}(\hat{\mu}_2 + \hat{\mu}_1)\Sigma^{-1}(\hat{\mu}_2 - \hat{\mu}_1) - \log(N_2/N_1),$$

and Class 1 otherwise.

2.2

Consider minimization of the least squares criterion

$$\sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2.$$

Show that the solution $\hat{\beta}$ satisfies

$$\left[(N-2)\hat{\Sigma} + N\hat{\Sigma}_B\right]\beta = N(\hat{\mu}_2 - \hat{\mu}_1),$$

where $\hat{\Sigma}_B = \frac{N_1 N_2}{N^2}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$.

2.3

Hence show that $\hat{\Sigma}_B\beta$ is in the direction $(\hat{\mu}_2 - \hat{\mu}_1)$ and thus

$$\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1).$$

Therefore the least-squares regression coefficient is identical to the LDA coefficient, up to a scalar multiple.

3 Problem 3

Build a kNN (k-Nearest Neighbor) classifier to predict wine quality using red wine quality data set (winequality-red.csv) from https://archive.ics.uci.edu/ml/datasets/wine+quality. Please do not use kNN related packages.

3.1

Plot 5-fold cross-validation errors on the same plot for each value in a sequence of discrete k's. What is the value of k selected by your cross-validation procedure? Comment on the shapes of the error curves.

3.2

Since the quality levels are ordinal, you can also use kNN regression by taking the average of the levels of the nearest neighbors and rounding it to the nearest level. Does it improve the classification?