

## SS9850B Assignment #2

(Due April 3 2020)

Please carefully read the following instructions:

- There are 3 questions in this assignment. 4 marks for each subquestion.
- Show all your works in details and provide your code for computations. Also, well summarize your numerical results in each question.
- This assignment is an INDIVIDUAL work. Plagiarism will earn ZERO mark.
- ONLY R functions/packages mentioned in this course are allowed. Using other functions/packages that are not used in this course will lose marks.
- Submit your solutions to the Drop Box in the course site. Delayed submission will lose marks.

1. **(Wine Data Set)** These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents (including Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, and Proline) found in each of the three types of wines. The sample size is 178. The dataset is available in the course site. The main interest of this dataset is to study multiclassification of the three types of wines. Let  $\hat{y}$  denote the predicted class of observations.

- (a) Instead of estimating  $\Sigma$  (or  $\Sigma_i$ ) by empirical estimates and then take an inverse, here we consider to use glasso to estimate  $\Theta \triangleq \Sigma^{-1}$  (or  $\Theta_i \triangleq \Sigma_i^{-1}$ ) directly. After that, replacing the estimator  $\hat{\Theta}$  (or  $\hat{\Theta}_i$ ) to the linear discriminant function (or quadratic discriminant function) gives *graphical-based* linear discriminant analysis (or quadratic discriminant analysis). In this question, use the graphical-based linear discriminant analysis and quadratic discriminant analysis to obtain  $\hat{y}$ . In addition, summarize the confusion table for  $y$  and  $\hat{y}$ , use macro averaged metrics to evaluate recall, precision, F-measure, and then conduct performance of classification.
- (b) Use the support vector machine method to obtain  $\hat{y}$ . In addition, summarize the confusion table for  $y$  and  $\hat{y}$ , use macro averaged metrics to evaluate recall, precision, F-measure, and then conduct performance of classification.
- (c) Suppose that we only consider the predictor **Proline**. Use the nonparametric density estimation method in Section 4.1 to explore the multiclassification. For the choices of kernel function, we examine the *Gaussian kernel* and the *biweight kernel*; for the bandwidth selection, we use the *cross-validation* (CV) method.

After obtaining  $\hat{y}$ , summarize the confusion table for  $y$  and  $\hat{y}$ , use macro averaged metrics to evaluate recall, precision, F-measure, and then conduct performance of classification.

*## Hint: programming code of CV function:*

```
J=function(h){
  fhat=Vectorize(function(x) density(X,from=x,to=x,n=1,bw=h)$y)
  fhati=Vectorize(function(i) density(X[-i],from=X[i],to=X[i],n=1,bw=h)$y)
  F=fhati(1:length(X))
  return(integrate(function(x) fhat(x)^2,-Inf,Inf)$value-2*mean(F))
}
```

- (d) Summarize your findings in (a)-(c).

2. **(Simulation studies)** Consider the following linear model:

$$y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_5\beta_5 - 5\sqrt{\rho}X_6\beta_6 + X_7\beta_7 + \epsilon, \quad (1)$$

where  $X = (X_1, \dots, X_p)$  is a  $p$ -dimensional vector of covariates and each  $X_k$  is generated from  $N(0, 1)$ . The correlations of all  $X_k$  except  $X_6$  are  $\rho$ , while  $X_6$  has the correlation  $\sqrt{\rho}$  with all other  $p - 1$  variables. Suppose that the sample size is  $n = 200$ .

- (a) Show that  $X_6$  is marginally independent of  $y$ .
- (b) Now, consider  $p = 1000$  and generate the artificial data based on model (1) for 1000 repetitions. Specifically, let  $\beta_i = 1$  for every  $i = 1, \dots, 7$  and set  $\rho = 0.8$ . After that, use the SIS and iterated SIS methods to do variable selection and estimate the parameters associated with selected covariates. Finally, summarize the estimator in the following table:

Table 1: Simulation result for (b)

	$\ \Delta\beta\ _1$	$\ \Delta\beta\ _2$	#S	#FN
SIS				
Iterated SIS				

- (c) Here we consider the scenario that is different from (b). Let  $p = 50$  and  $X \sim N(0, \Sigma_X)$  with entry  $(j, k)$  in  $\Sigma_X$  being  $0.6^{|j-k|}$  for  $j, k = 1, \dots, p$ . We generate the artificial data based on (1) for 1000 repetition with  $\beta_i = 1$  for every  $i = 1, \dots, 7$ . After that, use the lasso, adaptive lasso, and Elastic net (set  $\alpha = 0.5$ ) methods to estimate the parameters. Finally, summarize numerical results in the following table.

Table 2: Simulation result for (c)

	$\ \Delta\beta\ _1$	$\ \Delta\beta\ _2$	#S	#FN
lasso				
adaptive lasso				
Elastic net ( $\alpha = 0.5$ )				

- (d) Summarize your findings for parts (b) and (c), respectively.

**Note:** Let  $\hat{\beta}$  be the estimator, then  $\Delta\beta$  is defined as  $\Delta\beta = \hat{\beta} - \beta$  with the  $i$ th component being  $\hat{\beta}_i - \beta_i$ . Therefore,  $\|\Delta\beta\|_1$  and  $\|\Delta\beta\|_2$  are defined as

- $\|\Delta\beta\|_1 = \sum_{i=1}^p |\hat{\beta}_i - \beta_i|;$
- $\|\Delta\beta\|_2 = \sqrt{\sum_{i=1}^p (\hat{\beta}_i - \beta_i)^2}.$

3. In the class, we have discussed that the estimator of  $\Theta$  can be obtained by solving

$$\Theta^{-1} - S - \lambda\Phi = 0,$$

where  $S$  and  $\Psi$  are defined in the note.

- (a) Let  $W$  denote the working version of  $\Theta^{-1}$  such that  $W\Theta = I$ , where  $I$  is the identity matrix. Show that

$$W_{11}\beta - s_{12} - \lambda\psi_{12} = 0, \tag{2}$$

where  $\beta = -\frac{\theta_{12}}{\theta_{22}}$ , and  $\theta_{12}$ ,  $\theta_{22}$ ,  $s_{12}$ , and  $\psi_{12}$  are defined in the note.

- (b) Suppose that  $\hat{\beta}$  is obtained by solving (2). Please show that  $\hat{\theta}_{12} = -\hat{\beta} \cdot \hat{\theta}_{22}$  and  $\hat{\theta}_{22} = (w_{22} - w_{12}^\top W_{11}^{-1} w_{12})^{-1}$ . Also, interpret the meaning of  $\hat{\theta}_{12}$ .

**Hint: Regarding simulation studies with 1000 repetitions.**

In Question 2, you are asked to use simulation studies with 1000 repetitions to estimate the parameters. Specifically, based on the  $k$ th artificial data that are independently generated, you are able to obtain the estimator, denoted by  $\hat{\beta}^{(k)}$ . As a result, with 1000 repetitions, the final estimator is given by  $\hat{\beta} = \frac{1}{1000} \sum_{k=1}^{1000} \hat{\beta}^{(k)}$ .