

Math 583 HW 10 2023 Due Wednesday, Nov. 15.  
 Final: Wednesday, Dec. 13, 2:45-4:45.

**A)** Classification with a neural network produced the following confusion matrix on training data. Compute the AER =  $\sum_{i \neq j} n_{ij} / \sum_{i,j} n_{ij}$ . Class  $a$  = motorcycle,  $b$  = airplane,  $c$  = bike, and  $d$  = car where the neural network classifies images from the above four classes. Note that the AER = (total – diagonal total)/total where the total is the sum of the confusion matrix entries.

	truth			
predict	a	b	c	d
a	202	29	33	33
b	29	252	3	20
c	31	2	188	73
d	38	17	76	174

**B)** For a support vector machine (SVM) using a radial kernel,  $Z \in \{-1, 1\}$ , and as a “binary regression model” (using  $Z = 2Y - 1$  instead of  $Y$ ),  $Z \perp \mathbf{x} | [\beta_0 + \sum_{i=1}^n \alpha_i \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|_2^2)]$  where the  $n + 2$  parameters  $\beta_0, \alpha_i$ , and  $\gamma$  need to be estimated.

- What is the estimated sufficient predictor ESP?
  - What is  $\hat{Z}$  if  $ESP < 0$ ?
- (Hint: if  $\hat{Y}$  would be 0, then  $\hat{Z} = -1$ .)

**C)** This problem simulates the two sample HD test  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  versus  $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$  where the  $m$  out of  $n$  bootstrap is used with  $m_i \approx 2n_i/3$  for  $i = 1, 2$ . The term cov gives the observed proportion of 100 runs where  $H_0$  was rejected = power. When  $H_0$  is true want cov  $< 0.11$  for 100 runs. When  $H_0$  is false want large power (the closer to one the better).

- Copy and paste the  $R$  commands for this problem, where  $n_1 = n_2 = 100$ ,  $p = 100$ , and  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$  so  $H_0$  is true. Give the cov from the output.
- Copy and paste the  $R$  commands for this problem, where  $n_1 = n_2 = 100$ ,  $p = 100$ , and  $\boldsymbol{\mu}_1 = \mathbf{0}$  and  $\boldsymbol{\mu}_2 = 0.13 \mathbf{1}$  so  $H_0$  is false. Give the cov from the output.

**D)** This problem plots the p-values for a microarray data set for multiple testing. Most of the code came from Giraud (2022, pp. 214-215). The Golub data set is downloaded from Hastie’s website. The data set has 3051 rows for the  $m = 3051$  genes. The first 27 columns correspond to patients with leukemia of type “ALL.” The next 11 columns correspond to patients with leukemia of type “AML.” The first plot gets the 3051 p-values from a two sample t-test for each gene, and adds the line  $Y = pvalue = 0.05k/m$  corresponding to  $\alpha = 0.05$ . The second plot is an enlargement of the lower left corner of the 1st plot corresponding to the first 750 ordered p-values. The Benjamani-Hochberg procedure used  $R_{BH}$  with the first 645 p-values  $p_{(1)}, \dots, p_{(645)}$ . Copy and paste the  $R$  code into  $R$ .

You may need to right click on the command window and hit “Enter” to get the line for the second plot, but the comment # may save you the step.

**E) 4.10.**

This problem does lasso for binary regression for artificial data with  $n = 100$ ,  $p = 101$  and 5 active population nontrivial predictors. If  $SP = \alpha + \mathbf{x}^T \boldsymbol{\beta}$ , then the 100 nontrivial predictors are in  $\mathbf{x}$  and  $\boldsymbol{\beta} = (1, 1, 1, 1, 1, 0, \dots, 0)^T$ .

a) Copy and paste the source and library commands into *R*. Then copy and paste the commands for this part into *R*. Relaxed lasso = lasso variable selection gets the binary logistic regression model to the predictors corresponding to the nonzero lasso coefficients. Then the response plot is made. Include the plot in *Word*.

Does the step function track the logistic curve?

b) Copy and paste the commands for this part into *R*. These commands do MLR lasso, then the relaxed lasso gets the binary logistic regression model to the predictors corresponding to the nonzero lasso coefficients. Then the response plot is made. For this data set, one more predictor was used than that in a). Include the plot in *Word*.

Does the step function track the logistic curve?

c) Copy and paste the commands for this part into *R*. The commands for this part use MLR forward selection with EBIC, and only nontrivial predictor  $x_4$  was selected. Then the binary logistic regression is fit using this variable and the response plot is made. Include the plot in *Word*.

Is the plot in c) worse than the plots in a) and b)?

**F) 4.11.** This problem does lasso for Poisson regression for artificial data with  $n = 100$ ,  $p = 101$  and 5 active population nontrivial predictors. If  $SP = \alpha + \mathbf{x}^T \boldsymbol{\beta}$ , then the 100 nontrivial predictors are in  $\mathbf{x}$  and  $\boldsymbol{\beta} = (1, 1, 1, 1, 1, 0, \dots, 0)^T$ .

a) Copy and paste the source and library commands into *R*. Then copy and paste the commands for this part into *R*. Relaxed lasso = lasso variable selection gets the Poisson regression model to the predictors corresponding to the nonzero lasso coefficients. Then the response plot is made. Include the plot in *Word*. The horizontal line is  $\bar{Y}$  and the jagged curve is lowess which tracked the exponential curve well until  $ESP > 3$ . Lasso overfit using 26 variables instead of 5. (The plots are labeled a), b), c), and d). Plot a) is the response plot. Plot c) is the linear transform that should be useful for OPLS. Plot d) is the residual plot corresponding to plot c).)

b) Copy and paste the commands for this part into *R*. These commands do MLR lasso, then the relaxed lasso gets the Poisson regression model to the predictors corresponding to the nonzero lasso coefficients. Then the response plot is made. For this data set, 20 variables were used. Include the plot in *Word*.

c) Copy and paste the commands for this part into *R*. The commands for this part use MLR forward selection with EBIC, and only nontrivial predictor  $x_5$  was selected. Then the Poisson regression is fit using this variable and the response plot is made. Include the plot in *Word*.

If the Poisson regression model is good, we would like the vertical scale to be not more than 10 times the horizontal scale in the OD plot. (This happened in a) and b).) Is the vertical scale more than 10 times the horizontal scale in the OD plot for this model?