Name:

PubH 7475/8475/Stat 8931 Midterm Exam (Spring 2015)
(12:00–12:30pm, march 31, 20135 *[Total 40 points]*

1. For a classification problem, we tried a classifier with a tuning parameter controlling the model complexity on a training dataset and a test dataset.

   1a. Mark out which of the following plots is most likely showing the training error and test error curves. (4 pts)

   1b. In your chosen plot, mark out which corresponds to the training error and test error respectively. (2 pt)

   1c. In your chosen plot, mark out which part corresponds to over-fitting. (2 pt)

2. True or false: Cross-validation is used to obtain an almost unbiased estimate of test error.

3. Is each of the below statement true or false: (1 pts for each)

   3a. Linear discriminant analysis (LDA) assumes that the predictors in each class have a normal distribution.

   3b. If there are categorical predictors, then LDA cannot be applied. (1 pts)

   3c. Since quadratic discriminant analysis (QDA) is more general than LDA, QDA always performs better than LDA. (1 pts)

   3d. If the outcome is binary, then linear regression cannot be applied. (1 pts)

4. β_i is the regression coefficient for variable j in a logistic regression model. Consider the penalties on $\beta = (\beta_1, ..., \beta_k)'$:

$$P_1(\beta) = \sum_{i=1}^{k} |\beta_i|; \qquad P_2(\beta) = \sum_{i=1}^{k} |\beta_i|^2; \qquad P_3(\beta) = \sqrt{\sum_{i=1}^{k} |\beta_i|^2};$$

$$P_4(\beta) = \sum_{i \neq j} |\beta_i - \beta_j|; \qquad P_5(\beta) = \sum_{i \neq j} ||\beta_i| - |\beta_j||;$$

4a. Which of the above penalties can be used for direct variable selection? (2 pts)

4b. If we know a priori that all the variables are simultaneously either related or unrelated to the response variable, for the purpose of better variable selection, which penalty (penalties) is (are) most suitable? (1 pt)

4c. If we know a priori that all the coefficients are almost equal to each other, for the purpose of efficient estimation, which penalty (penalties) is (are) most suitable? (1 pts)

4d. What is a possible advantage of P_2 over P_1? (1 pt)

4e. True or false: The truncated Lasso penalty (TLP) covers the Lasso penalty as a special case. (1 pt)

4g. Compared TLP, what is the main disadvantages of P_1?

5. True or false: Partial least squares can be used when the number of predictors is much larger than the sample size. (1 pt)

6. True or false: Bagging classification trees often improves the performance of using a single classification tree. (1 pt)

7. True or false: Bagging classification trees often improves the performance of using a random forest. (1 pt)

8. True or false: Model averaging always performs better than model selection. (1 pt)

9. True or false: A fitted regression tree (CART) model is piece-wise constant. (2 pts)

11. True or false: AdaBoost can be formulated as a forward stagewise additive modeling. (2 pts)

12. What is the loss function (or its name) used in AdaBoost? (1 pt)

13. Is it true that SVM can be formulated as a penalized method? If true, give the corresponding loss function (or its name) and its penalty function (or name). (3 pts)

14. Explain briefly what is the kernel trick. (2 pts)

15. Is it true that K-means imposes less modeling assumptions than the Normal mixture model-based clustering? Why or why not. (2 pts)

16. Consider Normal mixture model-based clustering for two clusters. If each plot shows our prior knowledge on the two clusters, give the best covariance structures for each. (2 pts for each)