15.095, Midterm Exam

November 5, 2018

Instructions:

1. There are 100 points in this exam.
2. You have 90 minutes to complete the examination.
3. This exam is closed book and closed notes.
4. Please explain your work carefully for the short answer and proof questions.
5. Good luck!

Problem 1: True/False (40 points)

Indicate whether each statement listed below is True or False. If the statement is False, write one sentence explaining your answer. (2 points each)

1. The R^2 in linear regression is always positive.
2. A regression model with $R^2 = 0.8$ on the training data and $R^2 = 0.8$ on the testing data may be underfitting.
3. A regression model with sum-of-squared errors = 1,200 on the training data and sum-of-squared errors = 2,400 on the testing data is definitely overfitting.
4. The goal of median regression is to find the linear model which minimizes the median residual on the training data.
5. The randomness of each tree in Random Forests only comes from using different subsets of observations in the training set to build the tree.
6. When using a big M formulation for a mixed integer optimization problem, the value of M will not influence the speed of the algorithm.
7. For a cutting plane method applied to a minimization problem, the objective value of the optimization problem that we iteratively solve increases or stays the same when we add a cut.
8. The misclassification error of a classification tree for two classes that has zero splits is always 0.5.
9. The AUC of a classification tree for two classes that has zero splits is always 0.5.
10. A feedforward neural network for regression with n nodes in the first hidden layer with activation functions $f(x) = \mathbb{1}\{x \geq b\}$ can always be represented as a regression tree of depth that scales linearly with n.
11. Neural networks often have many hyperparameters (activation functions, number of layers, number of nodes per layer, etc.), but most of these do not need to be tuned.
12. The goal of cross-validation is to improve the performance of the model on training data.

13. Warm starts are used in mixed integer optimization to improve the computational time required to find an optimal solution.

14. The initial assignment of cluster centers impacts the final solution found by standard K-means clustering.

15. Two trees with the same misclassification objective value may have different gini impurity objective values.

16. Branch-and-bound is a method which is used to prove optimality for linear optimization problems.

17. The primary advantage of optimal classification and regressions trees over the usual approach is an increase in the models' predictive performance.

18. Robustness can be accomplished by adding regularization.

19. The Coefficient of Prescriptiveness can be negative.

20. Neural networks are provably more powerful than optimal regression trees.

Problem 2: Short Answer (30 points)

Briefly answer each question in 3-4 sentences. (6 points each)

(a) Discuss the tradeoffs between modeling power and interpretability for the (axis-parallel) Optimal Classification Trees and Optimal Classification Trees with hyperplane variable splits.

(b) Describe in words the quantity that is minimized in Optimal KNN imputation. How do we solve this problem in practice?

(c) How are good feasible solutions useful for the branch-and-bound method?

(d) Describe the bias-variance tradeoff in the context of classification trees.

(e) We are given historical data (x_i, y_i) from days $i = 1, \ldots, n$, where x_i are side data for day i and y_i is the demand for newspapers on day i. Please formulate the problem of maximizing the number of newspapers sold on a new day with side data x_{n+1} using a prediction/prescription framework with random forests.

Problem 3: Linear Regression (30 points)

Assume that we are given a data set for regression $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, n$, where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$, and we would like to predict y given $\mathbf{x}$. The traditional Least Squares (LS) estimator is:

$$\hat{\beta}^{LS} \in \operatorname*{argmin}_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2,$$

where $(y_i - \mathbf{x}_i^T \beta)^2, i = 1, \ldots, n$, are the squared residuals.

Please answer the following questions (6 points each). For parts (a)–(d), write out the full optimization formulations.

(a) Formulate the problem of minimizing the maximum value of the absolute value residuals $|y_i - \mathbf{x}_i^T \beta|,\ i = 1, \ldots, n$, as *linear* optimization problem.

(b) How can we modify this formulation so that it is robust to perturbations in the features?

(c) How can we modify this formulation so that it gives sparse solutions?

(d) Suppose that we know that the pairs of variables (1,3) and (7,8) are both pairwise collinear. How can we modify this formulation to ensure that our final model will not have pairwise collinearity?

(e) Please outline how you would solve the problem. Describe **two** ways that you could possibly speed up the solution time.