Final Exam 2020

CS 750/850 Machine Learning

The best 4 out of 5 problems count (25% each). Problem 5 is harder than others, so you may leave that one for the end. Turn your solution in online on mycourses. Scanned paper or typed are both OK. Even if you cannot solve a problem, please submit a partial solution showing your work to get partial credit.

Questions: Please email mpetrik@cs.unh.edu. I will hold virtual office hours on Monday 9:00 - 10:00, or we can arrange a zoom session over email.

Problem 1

Assume that you have a dataset with a predictor (feature) X and the target Y. You run simple linear regression ($Y \sim X$) and get the best fit with RSS = 20 and TSS = 120?

1. What is the *covariance* between X and Y if the variances of X and Y are $\text{Var}(X) = 15$ and $\text{Var}(Y) = 20$.
2. Repeat if RSS = 50 and TSS = 40. Discuss the result.

Problem 2

Suppose that we estimate the regression coefficients in a LASSO model by solving

$$\min_{\beta_0,\ldots,\beta_p} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \frac{1}{\lambda} \sum_{j=1}^{p} |\beta_j|$$

for a particular value of λ. For parts (1) through (5), indicate which of i. through v. is correct. **Justify** your answer.

1. As we increase λ from 0.1 to ∞, the training RSS will *typically*:

i. Remain constant.
ii. Steadily increase.
iii. Steadily decrease.
iv. Increase initially, and then eventually start decreasing in an inverted U shape.
v. Decrease initially, and then eventually start increasing in a U shape.

2. Repeat (1) for test MSE.
3. Repeat (1) for (squared) bias.
4. Repeat (1) for variance.
5. Repeat (1) for the irreducible error (Bayes error).
6. How would you choose the best value of λ?

Problem 3

1. Do you expect random forests to achieve a smaller or larger *training error* than bagged trees with the same number of trees? Justify your answer.

2. How would you expect the *training error* of bagged trees to compare with boosted trees? Justify your answer.
3. How would you decide how many trees to use for any particular dataset in a random forest? It is best to choose the number that minimizes the training error? Why or why not?

Problem 4

This problem examines the differences between SVC (linear SVM), SVM with a polynomial kernel, and other linear classifiers.

1. For an arbitrary training set, would you expect for SVC (linear) or SVM (polynomial kernel) to work better on the *training set*? Why?
2. If the Bayes decision boundary between the two classes is linear, would you expect SVC or SVM to work better on the *training set*?
3. *True or False*: There is no need to use slack variables in SVMs with polynomial kernels because the decision boundary can be nonlinear. Justify your answer.
4. LDA, Logistic regression, and SVC (linear) all fit a linear decision boundary. Will their fits be the same? What would be some reasons for you to prefer LDA over SVC?

Problem 5

1. Show that the method of *least absolute deviations* (https://en.wikipedia.org/wiki/Least_absolute_deviations) is the maximum likelihood solution to fitting a linear function to a target when assuming that the prediction error is distributed according to the Laplace distribution.
2. Is the method of *least absolute deviations* the same as or different from *median regression* (quantile regression with $\tau = \frac{1}{2}$: https://en.wikipedia.org/wiki/Quantile_regression). Prove it.