

83

ST 563 601 – SPRING 2025 – POST Final Exam Tablet

Student's Name: Ali Shashaaani

Date of Exam: Monday, April 28, 2025 - Wednesday, April 30, 2025

Time Limit: 90 minutes

Allowed Materials: None (closed book & closed notes)

Student – NC State University Pack Pledge

I, _____ have neither given nor received unauthorized aid on this exam or assignment. I have read the instructions and acknowledge that this is the correct exam.

STUDENT'S PRINTED NAME

STUDENT SIGNATURE



DATE

Exam must be turned in by:

EXAM END TIME

*STUDENT'S
INITIAL
AGREEMENT*

**NOTE: Failure to turn in exam
on time may result in penalties
at the instructor's discretion.**

Final Exam

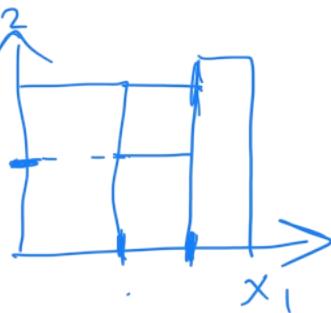
Please write your answers below each question. You should not have access nor use any materials during this exam.

A reminder that, by taking this exam, you are required to uphold the NC State honor pledge:

"I have neither given nor received unauthorized aid on this test or assignment."

1. We know that a multiple linear regression model fits a (hyper) plane as the response surface (or a curved hyperplane with higher order polynomial or interaction terms). How does a standard regression tree model the response surface?

Regression tree model Partitions the feature space into multiple regions. ✓
OK



2. For a standard regression tree that uses recursive binary splitting, suppose we have two predictors X_1 and X_2 . What criterion is used to determine the first split? a) Describe how this first split is decided upon. Be specific on both of these!

we use recursive binary splitting .

a) with greedy algorithm

meaning we go through all possible values of both X_1 and X_2 as a split point and we choose the

b) point that delivers the minimum RSS among all
Then we continue this method to have large tree

0

3. Suppose we have a large data set where we want to perform a regression task. We want to determine the best overall model between a kNN model and a ridge regression model. We want to use a train/test split and compare the best kNN and ridge regression model on the test set. We wish to determine the appropriate tuning parameters on the training set only using the bootstrap. Fully outline the process for splitting the data, tuning, comparing, and fitting a final overall best model.

- we start with train/test splitting of data (30/70) i.e.
- bootstrapping on the training set: bootstrap is a sampling method with replacement with 1.0 datapoints as the validation sets.
- Now we employ KNN (K is the tuning parameter)
nonparametric model and Ridge regression (Penalty factor is tunable - second order L^2)
- we tune our parameters using Bootstrap 10 validation set (Based on the error rate) ^{both} ^{not clear} - 2
- Finally fit 2 models with best tuned parameter to the test set and compare the results with error rates \rightarrow MSE, ...

At best to all data - 1

4. We discussed two ways to do 'early stopping' in a regression or classification tree. What are those two methods?

- number of splits ✓
- number of obs in terminal nodes

5. In a standard multilayer feed-forward neural network, what are two common activation functions?

- sigmoid ✓
- ReLU

6. What task is a Recurrent neural network well-suited for?

sequential tasks i.e. time series → stock market
for text classification → imdb reviews
improved model: bidirectional RNN

7. True or False questions (write True or false next to each letter):

a. Random forest and bagged tree models generally require you to standardize your predictors **False**

b. kNN models generally require you to standardize your predictors **True**

c. The number of trees we use in a random forest model is important because we can overfit with too many trees. **True**

d. When using BART we need to remove the first few prediction models. **True**

e. SVM models can only be used in classification tasks. **False**

f. KMeans clustering does not necessarily create the same clusters in each run of the algorithm. **True**

g. Hierarchical clustering requires you to know the 'true' underlying groupings to use it effectively. **False**

h. In a standard multilayer neural network, all inputs are 'connected to' all first level activations. **True**

i. KNN provides a discriminant for classifying our observations **False**

j. The Naive Bayes provides a discriminant for classifying our observations **True**

parametric

8. Consider the piecewise polynomial regression model. Here we define our knots to be c_1, \dots, c_M and use the indicator functions

$$h_1(X) = I(c_1 \leq X < c_2), \dots, h_{M-1} = I(c_{M-1} \leq X < c_M), h_M(X) = I(X > c_M)$$

in our regression equation given by

$$Y_i = \beta_0 + h_1(X_i)\beta_1 + \dots + h_M(X_i)\beta_M + \epsilon_i$$

Suppose we have n observations and we fit the model.

- a. What is the estimate of β_0 in this model?

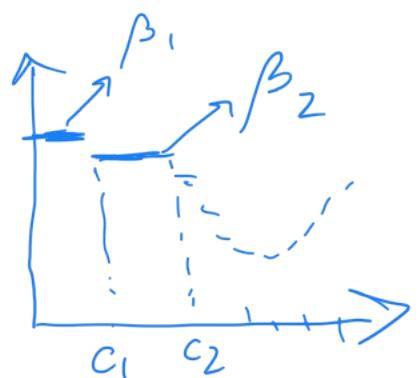
intercept

The value of y_i when $X_i < c_1$

$$h_0(x) = I(x < c_1)$$

∂c

No spline!



- b. What is the estimate of β_1 in the model?

β_1 is the parameter of the basis function $h_1(x)$
 when $c_1 \leq x \leq c_2$ and $h_1(x)$ is an indicator function
 therefore $y_i = \beta_1$ in that region. - |

9. What are the three most common tuning parameters associated with a boosted tree model?

b → iteration ✓

d → depth ✓

λ → shrinkage (learning rate) ✓

- |

10. Why do random forests for a regression task generally improve prediction over the basic bagged tree model?

basic bagged often suffers from correlation among
~~predictors~~ trees -1

RF use a random subset of predictors
↳ less correlated variables → better reduction in RSS
trees

11. Describe the algorithm for fitting a basic boosted regression tree model.

No SGD

sequential
nr boot strapping

we start from basic learner (weak learner)

then we sequentially fit multiple on what?
trees (shallow to complex) and every time -1

and in each sequence we slightly improve
model performance based on RSS value and
finally calculate the weighted sum of all.

Tunables: $B^{\text{iteration}}$, $d_{\text{tree depth}}$, λ shrinkage

12. When fitting a support vector machine model for classification, what are support vectors?

set of closest data points to

the margin (hyper plane)



13. When we wish to apply the SVM model to a classification task with more than two levels, we discussed the one-versus-one approach. Describe how this SVM model works.

Let's say we have 3 Classes levels, one vs. one approach is where we have one SVM for every possible classification between single levels.

1 vs 2 } other method is 1 vs all 1 vs 2, 3
1 vs 3 }
2 vs 3 }

then? - 1

14. Why do we often run the kmeans clustering algorithm multiple times?

14. Why do we often run the kmeans clustering algorithm multiple times?

K means clustering use random cluster assignment at the first step. so the results sometimes are not consistent enough. usually to control the variability we run the process multiple times and compare the results to choose the best one.

15. When doing hierarchical clustering, how does the 'single' linkage create a dissimilarity measure?

single linkage calculate the minimum distance among all 2 data points in 2 different clusters (intercluster)

- ## 16. What is a biplot and how can it be useful?

it is used for PCA where we compare 2 PC's together

There are data points available and we can see the weight of predictors over each PC and their direction

