

68

# ST 563 601 – SPRING 2025 – POST Final Exam Tablet

**Student's Name:** Chris Goodwin

**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, Chris Goodwin 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*

30 Apr 2025

*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?

With branches or splits at the predictors and the average response based on those splits. *ok*

The trees start with a root node for the first split and terminate at leaves.

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? Describe how this first split is decided upon. Be specific on both of these!

 You would plot  $X_1$  vs  $X_2$  and at each possible value of  $X_1$ , you would calculate the residual sum of squares (RSS) for each observation. You would choose the point with the smallest RSS to make your first split.

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.

1. First, split the data into a ~~training set~~ and a test set (e.g., 80/20 or 70/30)
2. For KNN you will create a grid of ~~tuning~~ parameters,  $K$ .
3. You will obtain  $B$  bootstrap resamples ~~from~~ and of the same size as the training data with replacement.
4. You will train a model for a value of  $K$  at each bootstrap resample ( $B$ ) and measure its performance by averaging the model metric for each (e.g. MSE) ~~on?~~ - |
5. Repeat for each  $K$  value.
6. Take the  $K$  for the best model based on the model metric.
7. For ridge you will create a grid of tuning parameters,  $\lambda_2$ .
8. Repeat steps 3 and 4, except for 4 you are using  $\lambda_2$ .
9. Repeat step 5 for all  $\lambda_2$
10. Take the  $\lambda_2$  for the best model based on the model metric.
11. Test Both the final KNN and regression model on the test set and measure performance using the model metric ~~Fit~~
12. ~~Test~~ the best model on the entire dataset.

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

Decide on depth of tree ✓

Decide on the number of observations at each leaf.

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

Filter layer

- 2

Pooling layer

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

Image recognition

- 3

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 False

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

- 9

- 14

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  $\beta_0$  in this model?

$\bar{Y}$

Its the value of  $Y_i$  when  $X < c_1$

- |

- b. What is the estimate of  $\beta_1$  in the model?

$\bar{Y}$

Its the increase in the value of  $Y_i$   
when  $c_1 \leq X < c_2$ ,  $x$  is greater than  
or equal to  $c_1$  but less than  $c_2$ .

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

- number of trees (sequential iterations) ✓
- number of parameters to use
- number of observations from original data set.

-2

-3

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

You take a random sample of the ~~observations~~ at each fitted tree, which reduces variance and prevents overfitting.

- 4

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

- You fit a basic regression tree model
- You use this tree to improve the model for the next tree
- With each sequential iteration, you learn from the previous tree model to improve the next tree model.

- |

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

They are the points closest to the decision boundary ✓ on which the performance of the decision boundary can be determined.

- 5

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.

You're comparing one cluster to one cluster at a time to determine the decision boundary.

- 2

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

With each iteration of the algorithm the clusters become better defined (based on distance measurements), when there is no more improvement in the average distance between the observations in each cluster, the algorithm can stop.

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

In a single linkage the smallest pairwise dissimilarity between the points in each cluster to determine the dissimilarity.

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

A biplot plots two predictors or features against each other and can be useful in determining decision boundaries for classification.

- 5