

54

ST 563 601 – SPRING 2025 – POST Final Exam Tablet

Student's Name:

Constantino Acptis

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, Constantino Acptis 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



4/30/2025

DATE

Exam must be turned in by:

EXAM END TIME

STUDENT'S

INITIAL

AGREEMENT

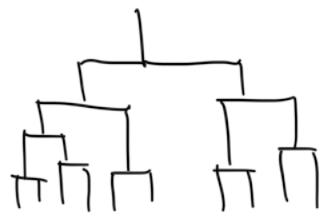
**NOTE: Failure to turn in exam
on time may result in penalties
at the instructor's discretion.**

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?



Trees uses recursive binary splitting to split the tree in branches, as we can see in the left. CP as tuning parameter.
Trees are non-parametric models.

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!

If it is randomly assigned. We would use the CP which is the tuning parameter to find the best way to split the tree.

- 5

- 7

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 by splitting the data into training and testing about (70%), (30%) on (80%, 20%). For the kNN we would tune the parameter k on training data, using ~~Bootstrap~~ ^{now?} then find k closest neighbor, using the Euclidean distance. Fit the model on the ~~test~~ set and repeat K times. For the ridge regression model, we tune the parameter λ on training data using bootstrap. Fit the model on ~~test~~ set and repeat n times. After, we would compare both models and choose the best model by the lowest MSE. We then fit the best model into the full dataset. -3

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

Early stopping means that it automatically stops when the model stops improving. CV and log-loss

-4

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

I'm confused if here I should talk about sign() function on the Filter layer and Pooling layer.

-2

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

RNN → Forward (Past \rightarrow Future) / If needs the ok
Backward (Future \rightarrow Past) / direction and has two directions

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

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

b. T kNN models generally require you to standardize your predictors

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

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

e. F SVM models can only be used in classification tasks.

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

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

h. T In a standard multilayer neural network, all inputs are 'connected to' all first level activations.

i. F KNN provides a discriminant for classifying our observations

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

-9

-15

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?

β_0 is the mean response for the first knot. - | ?

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

β_1 is the mean response for the interval $c_1 \leq X < c_2$. ok

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

- Expanding the tree and change the preds in the leafs.
- Pruning which is cutting branches of the tree
- Growing branches - 3

* reducing the bias and resulting in a better prediction. A single tree can be underfitted and not capture complex patterns in the data. -4

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

When we talk about random forests, we have a lot of trees compared to ~~one~~ basic bagged tree. We need to be careful to not overfit the training data with that many trees. This will allow for a higher flexibility.

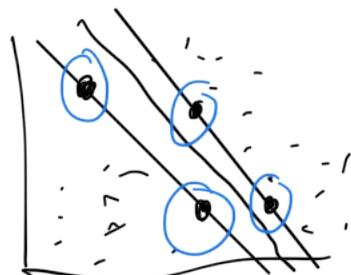
11. Describe the algorithm for fitting a basic boosted regression tree model.

Each tree \rightarrow randomly sampling with replacement, (as the size of the original).
on choosing data points some would be OOB.
we would test on unseen data such as the OOB.
And the final prediction = Avg of prediction for all trees, this for Random forest.

-5

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

Support vectors are the points where the margin is formed, and they "support" those lines for the margin.



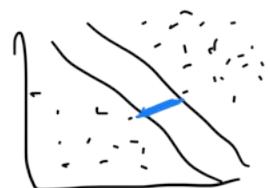
F9

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.

- SVM model for each class vs ~~the others~~.
- Each model will give a score
- choose the highest score. -3

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

Because KMeans clustering does not create the same clusters in each run of the algorithm.
It's also very sensible for changes depending on the amount of points. -1



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

Single:

, smallest distance between pair of ~~up~~ points on each side.

Complete would be the maximum and AVG would be the mean of all points.

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

-3

Biplots allow us to analyze multiple variables within the same plot.

-7