39

ST 563 601 – SPRING 2025 – POST
Final Exam Tablet

Student's Name: Koji Takagi

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, Koji Takagi 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: 9:00 *Exam End Time* K.T. *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?

Classification and regression tree's response surface is simple line. -3

2. For a standard regression tree that uses recursive binary splitting, suppose we have two predictors X1 and X2. What criterion is used to determine the first split? Describe how this first split is decided upon. Be specific on both of these!

The criterion is binary (make a cut off for the question) and use majority rule. We choose most common classification at the leaf node. -0

-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.

train/test
-1

First, using bootscrape sampling with replacement, (training data), out of bag is the test data.

We create the model using training data, and ~~fit~~ the model using the test data.

tuning grid
-1

When creating the model, we decide tuning (k for KNN, weight for parameter (ridge)) using training data

Bootstrap done repeatedly -2

When comparing the kNN and ridge regression, we compare AUC or BIC of those -2 to choose the overall best model.

-6

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

Boosting tend to overfit, so we use 'early stopping' methods.
1. We decide when to stop in the bagging. - 4
2. We use cut off to stop.

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

Filter: extract features (like edge)
Pooling: reduce size (max pooling) -2

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

time series or text
(sequentials) ✓

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. ~~false~~
   e. SVM models can only be used in classification tasks. ~~true~~
   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. ~~false~~
   i. KNN provides a discriminant for classifying our observations ~~true~~
   j. The Naive Bayes provides a discriminant for classifying our observations ~~true~~

-2

-27

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

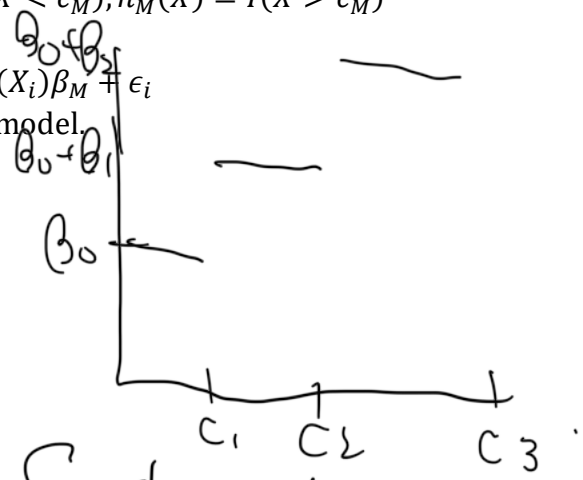
$h_1(X) = I(c_1 \leq X < c_2), \ldots, h_{M-1} = I(c_{M-1} \leq X < c_M), h_M(X) = I(X > c_M)$

in our regression equation given by

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

Suppose we have n observations and we fit the model.

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



When $X < c_1$,

$Y_i = \beta_0$

β_0 is the baseline for the model.

Y_i shows β_0 when $X_i < c_1$.

-1

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

only when $X_i (c_1 \leq X < c_2)$,

ok

β_1 is the increase from β_0 ✓

Y_i shows $\beta_0 + \beta_1$ when $X_i (c_1 \leq X_i < c_2)$

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

- early stopping
- regularization
- support different loss function

-3

-4

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

Not only bagging, but the Random forests also select random variables in each split. So, in addition to bagging, random forests can improve prediction.

-3

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

Boosting creates trees one-by-one.
The tree tries to fix the previous residuals.

-3

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

Support vectors are points closest to the margin.
It helps with determing a boundary.

-6

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.

one-versus-others. if we have three levels (A, B, C), we compare A vs (B, C), B vs (A, C), C vs (A, B), then check highest score to decide classification. -3

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

First, K is determined randomly .. and assign the points close to the centroid, then minimized the within cluster distance. -3
K and centroids are randomly selected. so we have to adjust to make better models. There fore, we often run the Kmeans multiple times.

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

We compare the distance of the points which are the closest across the cluster ✓
Please see the figure →

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

When we choose how many dimensions to use the model, biplot can show the image -5
We should choose the numer of dimensions.

-11