

72

ST 563 601 – SPRING 2025 – POST Final Exam Tablet

Student's Name:

Jarrett Glass

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, Jarrett Glass

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

28 Apr 2025

DATE

Exam must be turned in by:

6:25 PM
EXAM END TIME

JG
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?

The regression tree models the surface by establishing the point at which the probability of a value going into one leaf or another is 0.5 - in other words, each node that determines where a value's placed is a "hyperplane" of sort.

- 1

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 recursion indicates that downstream data + decisions will have an impact on earlier splits as well. The first split in a tree is determined by minimizing error rate and accuracy based on misclassification factors (Gini index, cross-entropy, etc), and provides the node that minimizes them.

- 5

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.

- Split the data by randomly selecting observations from the full data set up to a partitioning percentage, usually as an 80/20 or 70/30 split. The training data set contains the larger amount. Will usually use the `createDataPartition()` function for this.
- Tune these by creating grids of the hyperparameters ("k" for kNN, " λ " for ridge regression) and evaluating the training data set against the hyperparameters for each value. That was worded well, but what this means is that for a grid of possible values of k (1 through 10), bootstrapping would be done to evaluate the effectiveness of the model when $k=1$ (using out-of-bag samples or testing), again when $k=2$, $k=3$, etc. up through $k=10$. The "best" MSE with the lowest value dictates the optimum k . Some process for establishing optimum λ .
- With optimum $k + \lambda$ values determined by bootstrapping from the training data, the models fit to? should be run against the TEST data that we initially separated. Which model has lowest test MSE between kNN or Ridge regression (using optimum)

\times or \checkmark) should be the preferred model and
fit to entire data -/→

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

1. specifying a tree depth ✓
2. Specifying a minimum number of observations per node ✓

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

1. ReLU ✓
2. Sigmoid

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

language learning models, speech recognition, and word completion.

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~~ ✓ (burn-in)
- 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, exploratory analysis is the whole point!~~
- 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~~

6

~~True~~

—6

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?

The average value of Y_i when $X_i < c_1$
(all other indicators $h(X_i) = 0$.) ✓

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

The average value of Y_i when $c_1 \leq X_i \leq c_2$.
ok

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

- tree depth ✓
- observations per node - |
- number of ~~bootstraps~~ iterations - |

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

Because the elements of randomness prevent overfitting directly to the training data.

- 4

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

Similar to bootstrapping, multiple trees would be generated on ~~subsets~~ of the data drawn from sampling with replacement. Trees are generated based on "useful features" from each subset (decision boundaries for each feature, tree depth & other hyperparameters). The final model is established as an average of those tree models.

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

The datapoints closest to the hyperplane that define the margin of it. ✓

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.

A separate binary model is created for each level, pitting that level against ~~all others~~ pooled together. The prediction is then run through each model, and a value is returned based on its distance from the hyperplane. The level's model yielding the largest value is the one assigned to the prediction. -3

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

Because the specific clusters may differ on each run, so running it multiple times allows one to take an "average" of the runs. -2

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

By copying the minimum of ~~distances~~ between two data points across clusters.

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

A two-dimensional model that has been projected to a higher-dimensional one to ascertain a clear hyperplane in the data. -1