DELTA Testing Services

go.ncsu.edu/testing

Campus Box 7555
1730 Varsity Dr.
Venture IV, Suite 236
Raleigh, NC 27695-7113

919.515.1560 phone
919.515.7180 fax

delta-testing@ncsu.edu

NC STATE UNIVERSITY

78

DELTA Testing Services

Student Name: Matthieu Cartron **Date:** 4/29/25

Student's NCSU Email Address: mccartm@ncsu.edu

Course: ST 563 601 **Exam #:** Final

Start Time: 1033 am **EndTime:** 1154 am

Proctor's Name (Print): Bilha Lucero

Proctor's Signature: Bilha Lucero

Institution: Univ. of New Mexico Testing & Training Ctr.

PLEASE SIGN & DATE THIS SHEET AND RETURN ALONG WITH THE EXAM

Proctoring Guidelines

If you are unable to comply with the following, please destroy the exam and have the student submit the name of another proctor for approval.

1. Please ask student for their photo ID.
2. **Have the student put their name on the exam and exam answer sheet.**
3. The test should be conducted in an atmosphere conducive to good concentration (quiet, good lighting, etc.).
4. The student must take the exam without outside help. Have the students leave all materials (except blank paper, pen or pencil, or calculator, as needed) outside the testing room. This includes notes, books, calculators, phones, etc. (excluding materials required for the exam).
5. Close and constant supervision must be provided.
6. Please scan and email the proctoring form, completed exam, and any formula sheets permitted for the assessment to delta-testing@ncsu.edu or fax to 919-515-7180.
7. Not including exams that permit all notes or textbooks, students should not be permitted to leave the testing room with formula sheets or scrap paper unless explicitly stated.
8. DO NOT GIVE THE EXAM TO THE STUDENT TO MAIL BACK

If you have any questions, please contact DELTA Testing Services at our main Venture IV location via phone: (919)-515-1560 or e-mail: delta-testing@ncsu.edu.

Thank you for assisting our students.
DELTA Testing Services
NC State University

Updated March 2022

ST 563 601 – SPRING 2025 – POST
Final Exam Tablet

Student's Name: Matthieu Cartron

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, Matthieu Cartron 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 4/29/2025 *Date*

Exam must be turned in by: 1203pm *Exam End Time* MC *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?

easier to see, but still applies with more splits

The splits created systematically divide up the predictor space (visually this will look like a grid of "regions" if we just have a few splits). In essence this will create a nonlinear surface through the predictor space.

prediction? -2

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!

With recursive binary splitting, the first split is done on the feature that furthest minimizes the sum of squared errors. For the first split, the entire feature space is explored (all of X_{1i} and X_{2j} are considered for this first split - one of these values from one of these features will yield the greatest reduction of sum of squared errors.

both X_1 & X_2

of ...

-2

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. Split the data into training and test groups, and initialize a tuning grid of values for k and λ (for the KNN model and ridge regression model, respectively)

2. Choose a number of bootstrap resamples to perform. For each value of k and lambda, fit the models on the resampled values, and evaluate this on the out of sample observations. Record the performance of each model with each tuning parameter on each out of sample validation set. For each model and each tuning parameter, take the average recorded performance (metric like RMSE) of each out of sample validation sets to get a single value for each candidate model.

for each bootstrap sample

3. For the KNN model, choose the value of k that produced the lowest average RMSE. Do the same for the ridge regression model, though here we want to choose the value of lambda that resulted in the smallest average RMSE

4. Refit our final KNN model and ridge regression model to the entire training set.

5. Evaluate how these two models perform on the test set. Choose the model that had the lower RMSE value. This is the final, "best" model.

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

- we can determine the maximum tree depth (or tune this)
- we can specify leaf size, that is, how many observations there must be (minimum) in the terminal nodes (leaves)

Can tune as well →

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

- RELU
- Sigmoid (logit)

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

Sequential data, e.g natural language processing

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 (thinning)

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

-3

-3

-6

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?

β_0 is the intercept - it does not have an associated basis function. Will exist regardless of the value of X.

-2

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

regression

If $c_1 \leq X \leq c_2$, β_1, the first spline coefficient, will exist (its spline basis function is the indicator with the condition that X lie between knots c_1 and c_2).

-1

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

- # of trees ✓
- tree depth (max) ✓
- learning rate ✓

-3

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

Random forests are an extension of bagged trees but instead of using all features, only a subset of the features is used to fit each tree (m). This prevents some important features from dominating the fit, and reduces the correlation between trees. This has the overall effect of reducing model variance and bias, which is why rfs generally outperform basic bagging.

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

With boosted regression trees, individual trees are fit sequentially on the residuals of the preceding trees, slowly improving fit. We start with an initialized tree, then sequentially iterate upon it (fitting on the residuals) until additional trees do not meaningfully improve the overall model fit. Fitting sequential trees has the effect of reducing bias and variance (if not overfit), making this method one that will often outperform standard regression trees and bagged trees.

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

The support vectors are the transformed features that allow us to "flatten" the feature space for the hyperplane fitting. The support vectors allow us to transform the data such that our data becomes separable.

-2

-2

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.

In the one versus one approach we pit each class against the other (all combinations): (3 different SVMs in this example)

A vs B -> -1, 1
A vs C -> -1, 1
B vs C -> -1, 1

The deciding vote for a given class is determined by the most probable (highest probability) class across each model

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

-1

The clusters resulting from this algorithm can change with each run, so we would want to run the algorithm numerous times to see how the clusters appear to generally be constructed. Running it only once might result in a cluster that the algorithm does not frequently produce (outlier).

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

With single linkage, dissimilarity is determined only by the distances between the final clustered observations.

-3

? final ?

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

Useful for visualizing clusters across two features.

-5

-9