ST 563 601 – SPRING 2025 – POST
Exam #1

Student's Name: Ali shashaani

Date of Exam: Thursday, February 6, 2025 - Friday, February 7, 2025
Time Limit: 75 minutes
Allowed Materials: None (closed book & closed notes)

Student – NC State University Pack Pledge

I, Ali shashaani 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 2.7.2025
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.

Exam 1

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. In the statistical learning paradigm, we discussed three major goals: statistical inference, predictive modeling, and pattern finding.

   Give a brief real world example for each of these goals. Specify a possible model or method we discussed in class that would help answer the question from each real world example.

- Statistical Inference (4 pts) $E(Y|X_i)$

This is how we can verbalize aur fitted model, explaning the relationship between aur predictors and response in suppervised learning.

- Predictive Modeling (4 pts) $\hat{Y}$

In predictive madeling we can predict the value af aur respanse variable by inputting requested (or arbitrary) predictive values in supervised learning

- Pattern Finding (4 pts)

In Unsupervised learning we dan't have a respanse variable. Insted we are looking for a pattern between explanatory variables

2. Consider having models characterized by flexibility with the scale going from not very flexible to very flexible.

   a. What type of relationship between flexibilty and squared bias would we expect? Why? (4 pts)

The more flexible the model is the more it is prone to over fitting which increases the (squared) Bias

   b. What type of relationship between flexibilty and variance would we expect? Why? (4 pts)

Models with higher flexibility tend to fit more observations to the model, so it increases the variance   ↑flexible & ↑ δ^2

   c. What type of relationship between flexibilty and training error would we expect? Why? (4 pts)

More flexible model tend to reduce training error as it tries to fit every observation inside the training set.

   d. What type of relationship between flexibilty and test error would we expect? Why? (4 pts)

we can not provide a general statement, but usually when we increase model's flexibility, since we stick to training data, we are possibly making test error larger

3. What is a tuning parameter or hyperparameter? How does this differ from a 'regular' parameter in a parametric model? (4 pts)

Tuning parameter, is a parater we use to construct our model based on, It tend to define our model structure. In parametric models: regular parameters are characteristics like mean and $\hat{\beta}$, and the tuning parameter is the size of our model

OLS not unique

MLR

5. In the multiple linear regression setting, we discussed a number of model selection methods. State four model selection methods that can be used in the $p > n$ situation. (4 pts)

shrinkage → LASSO

combining parameters → principal component reg

reduction methods → best subset selection, forward, Backward

6. State true or false (no need to explain). (3 pts each)

   a. Ordinary least squares performs variable selection. False

   it is possible by methods like forward step, best subset and backward step.

   b. Ordinary least squares performs shrinkage of coefficient estimates. False

   c. Best subset selection performs variable selection. True

   d. Best subset selection performs shrinkage of coefficient estimates. False

   e. Ridge Regression performs variable selection. False

   f. Ridge Regression performs shrinkage of coefficient estimates. True

   g. LASSO performs variable selection. True

   h. LASSO performs shrinkage of coefficient estimates. True

7. 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 LASSO model and a kNN regression model. We want to use a train test split and compare the best kNN and LASSO model on the test set. We wish to determine the appropriate tuning parameters on the training set only using cross-validation. Fully outline the process for splitting the data, tuning, comparing, and fitting a final overall best model. (10 pts)

- we initially perform data splitting to training and test sets, using v fold CV.
- for kNN we set "K" as our tuning parameter, we define a grid for values of k, we fit our model with each k value in training folds and test it by measuring test error on the validation fold. then we combine the test errors. we repeat that process for all values of k and then we compare test errors to select optimal k with least test error.
- For Lasso the process is similar except our tuning parameter which is λ. we finally select optimal λ based on the test error comparison between

different values of λ.

- The final step is to fit our kNN and Lasso model on the test set. with their optimal tuning parameters, and we select the model with the least test error.

8. Consider the Ridge Regression procedure for fitting a multiple linear regression model. With this model we minimize the following criterion (recall $\lambda \geq 0$):

$$\sum_i \left(Y_i - \beta_0 - X_{i1}\beta_1 - \cdots - X_{ip}\beta_p\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

a. What are the benefits of fitting a Ridge Regression model as compared to an ordinary least squares model? (4 pts)

Ridge reg is a L2 shrinkage model, it shrinks our $\hat{\beta}$ towards zero (not equal to zero) in order to reduce flexibility of the model and optimize variance (reduction) and bias (increase) in compare to OLS

b. What happens to our coefficient estimates for a 'large' value of the tuning parameter? What happens for a tuning parameter value near 0? (4 pts)

$\lambda \simeq 0 \rightarrow$ The model is similar to OLS

λ Large $\rightarrow$ The model shrink too much and undefitting is possible we may increase bias (squared)

9. Suppose we fit a multiple linear regression model to data about how much people earn. Our response variable is the wage (in 1000's of dollars) and our predictors are `marital_status` (`married`, `never_married`, or `divorced`), and age.

   We fit a linear and quadratic term for age and include an interaction between `marital_status` and age and an interaction between `marital_status` and age squared in the model. Output for the model is given below.

		Estimate	Std. Error	t value	Pr(>|t|)
	(Intercept)	25.293	38.116	0.664	0.507
x_1	marital_statusmarried	-19.780	40.405	-0.490	0.624
x_2	marital_statusnever_married	-31.760	40.992	-0.775	0.439
x_3	age	2.846	1.611	1.767	0.077
	I(age^2)	-0.024	0.017	-1.470	0.142
	marital_statusmarried:age	2.024	1.716	1.179	0.238
	marital_statusnever_married:age	2.230	1.820	1.225	0.221
	marital_statusmarried:I(age^2)	-0.025	0.018	-1.412	0.158
	marital_statusnever_married:I(age^2)	-0.032	0.020	-1.607	0.108

   a. Write down the fitted equation for $\hat{y}$. Define any indicator variables as needed. (4 pts)

$\hat{y} = 25.29 - 19.78$ martial-married $- 31.76$ nevermarried
$+ 2$ age $- 0.02$ age^2 $+ 2.02$ married.age $+2$
2.23 nevermarried-age $- 0.02$ marrild age^2 $- 0.03$ never marriedage^2

   b. One column of the output represents the t-value or t-statistic. What is the usefulness of this t-value? (2 pts)

t-value $= \dfrac{\hat{\beta}_i - 0}{SE}$ to be used for p-value
based on the HT of $\hat{\beta}_i = 0$ & H_0 in the
fitted model

$\hat{y} = 25.29 - 19.78 + 2 \times 30 - 0.02 \times (30)^2 + 2.02 \times 30$
$- 0.02 (30)^2$

c. Write down the form of a predicted value for somone that is `married` and has an age of 30. No need to simplify. (2 pts)

d. Write down the form of a predicted value for somone that is `divorced` and has an age of 30. No need to simplify. (2 pts)

$\hat{y} = 25.39 + 2 \times 30 - 0.02 (30)^2$

f. Conceptually, what does including an interaction between `marital_status` and age and an interaction between `marital_status` and age squared do to our model as compared to a model without those interactions (that still includes a main effect for `marital_status` and a linear and quadratic term for age)? (3 pts)

It changes the slope for age and age^2

$F(8, 2991) = 46.2$, p-value ≈ 0

g. The F-statistic for the global model test is 46.26 on 8 numerator and 2991 denominator degrees of freedom. The p-value for the test is very close to zero.

i. Write down the null and alternative hypotheses for this global test. (3 pts)

$H_0: \beta_i = 0$
H_a: at least one β is not zero

ii. We see a significant global test but none of the coefficient tests are significant. What do you think could be causing this issue? (3 pts)

multicalinearity

h. What type of plot might we look at to investigate the homogenous error variance (i.e. the assumption of equal error variance)? (3 pts)

residual plot
QQplot to check normality