ST 563 601 – SPRING 2025 – POST
Exam #1

Student's Name: Nirmal Timilsina

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, NT 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

NT 2/7/2025

Student Signature *Date*

Exam must be turned in by: 2:45 NT

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)

→ Statistical inference is used to find the importance and relationship of the predictor in the model. If we want to know the relationship between salary based on gender & education, we can use statistical inference to get the relationship & their importance (gender or education). We can use linear regression model.

- Predictive Modeling (4 pts)

→ Predicting modeling is used to predict the future value. In real world, Suppose we want to predict plant biomass based on nutrient, temperature and soil moisture. We can use linear regression, decision trees or support vector machines that has best predictive power based on data to predict for future & improve management decisions.

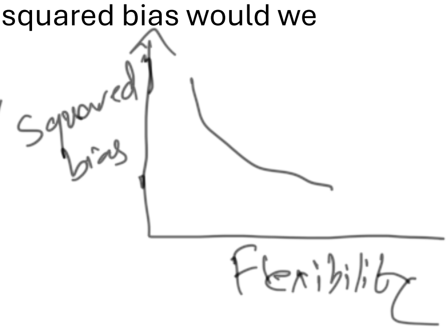
- Pattern Finding (4 pts)

⇒ We can use statistical methods for pattern finding for datas with unlabelled data. We can cluster based on their characteristics to find patterns, which helps statistical methods to learn & train modes. Cluster analysis can be used for pattern finding.

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

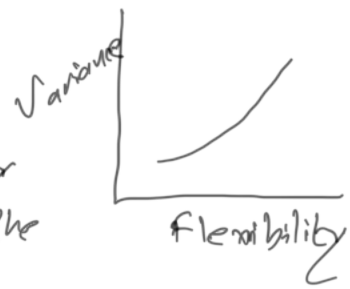
   a. What type of relationship between flexibilty and squared bias would we expect? Why? (4 pts)

→ squared bias decreases with increasing model flexibility. Flexible models can learn data and fit more accurately, thereby decreasing bias.



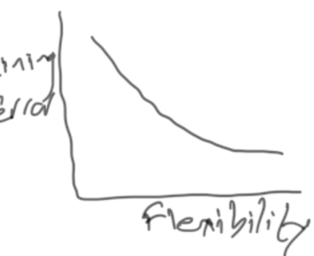
   b. What type of relationship between flexibilty and variance would we expect? Why? (4 pts)

→ Variance increases as model flexibility increases. When the model flexibility increases, the model learns the data & overfits it as flexibility increases, thereby increasing variance. Variance is the variability of a model in predictions.



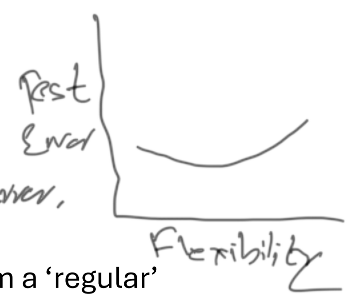
   c. What type of relationship between flexibilty and training error would we expect? Why? (4 pts)

→ Training Error decreases as flexibility increases. When flexibility increases squared bias decreases, thereby decreasing training error.



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

→ When flexibility increases, test error decreases at first. But when model becomes more & more flexible, it starts overfitting & data, thereby increasing test error.



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

→

Parameter	Hyperparameter / Tuning parameter
- Parameter are estimated from the data.	- Hyperparameter are values picked that tunes/optimizes the model.
- In parametric model, parameters are important and are structured.	- Tuning parameters are optional & helps improve the model

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)

→ The model selection methods are:
- lasso regression
- Ridge regression
- Partial least squares
- Elastic net

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

   a. Ordinary least squares performs variable selection.

   → False

   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)

→ To perform a regression task & compare & find best model between LASSO & kNN model we can follow the process below:

A> Split the data into ~~traing~~ training & test set (80:20 or 70:30 or 60:40 traing data : test data)

B> Tuning the hyperparameters using V-fold cross validation.
  a> Split the training set again into v-fold of equal sizes.
  b> Train the model on v-1 fold for each tuning parameter & used the remaing fold to get test metrics (MSE, RMSE, R^2) on validation set.
  c> repeat the process a & b and average the test metrics. (MSE, RMSE, R^2)
  d> Choose the best tuning parameter. for each reg. methods.

C> Fit the models on test set with best tuning parameter and calculate the test metrics (MSE, RMSE, R^2)

D> Compare the metrics between the model & select the best method. between Lasso & KNN.

E> Fit the best method to overall data & predict the response.

Note: For kNN, diff k are tuning parameters & for lasso. λ can be used with diff. values.

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

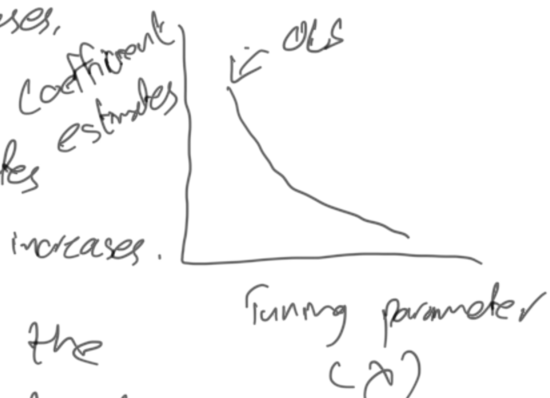
$$\sum_i \left(Y_i - \beta_0 - X_{i1}\beta_1 - \cdots - X_{ip}\beta_p\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

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

→ Ridge regression introduces the penalty term to the model. This regularizes & shrinks the parameters. This simplifies the model by making it more restrictive. The model is more interpretable. Ridge regression can also be fitted when predictors are larger than observation, when least squares cannot be fitted.

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)

→ As tuning parameter value (λ) increases, the model becomes less flexible & more restrictive. The coefficient estimates decreases when tuning parameter (λ) increases.



At tuning parameter value near 0, the model is near or equal to ordinary least squares model. The coefficient are not shrinked or shrinked by a very low.

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
marital_statusmarried	-19.780	40.405	-0.490	0.624
marital_statusnever_married	-31.760	40.992	-0.775	0.439
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)

Salary $(\hat{y}) = 25.293 - 19.780 * (\text{married}) - 31.760 * (\text{never married}) + 2.846 * \text{age} - 0.024 * (\text{age}^2) + 2.024 * (\text{married : age}) + 2.230 * (\text{never\_married} * \text{age}) - 0.025 * (\text{married} * \text{age}^2) - 0.032 * (\text{never-married} * \text{age}^2)$

   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 or t-statistics can be used to determine the importance of the corresponding predictor in the model. It is also used to calculate p value for hypothesis testing & finding if is an important predictor in the model.

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)

$\hat{Y} = 25.293 - 19.780 + 2.846 * 30 - 0.024 * (30^2) + 2.024 (1 * 30) - 0.025 * (1 * 30^2)$

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)

$\hat{Y} = 25.293 + 2.846 * 30 - 0.024 * (30^2)$

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)

→ Including the interaction hepls us fit the model with different slope. Includi interaction also helps when we need to fit non-linear data with linear models. They also provides importance when finding importance of predictor. If interaction is significant, both predictors are regardless of main effect.

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)

Null hypothesis: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ (beta coeffts. are zero)
Alternate: $\beta_1 = \beta_2 = \ldots = \beta_p \neq 0$ (atleast one coeff. is not zero)
Interpretation: If p is very close to zero (<0.05), We reject null hypothesis.

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)

→ Global test significant means the model is significant. If none of the coefficient tests are significant, then it could mean the the model assumptions are not properly met. There might be variability in data or has outlier or influencial points.

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)

→ To investigate the homogenous error variance, we can fit the residual versus fitted plot. The errors should be distributed in the plot without making any patterns or shapes to meet the condition of equal error variance.