MA 506 Probability and Statistical Inference

Instructor: Prashant Shekhar, PhD

Tentative Schedule for Fall 2022

Week Number: Starting Date (days)	Topic	Homework	Learning Outcome
Unit I: Course Basics			
1: 29th Aug (M,W,F)	Course introduction Python basics Python basics		1 1,2 1,2
2: 5th Sept (W,F)	Probability Statistical inference		1,2,16 1,2,16
Unit II: Regression Inference			
3: 12th Sept (M,W,F)	Regression I Regression II Regression III	HW1 released	4,5,6,7 4,5,6,7 4,5,6,7
4: 19th Sept (M,W,F)	Linear Regression from scratch Regression diagnostics Confidence Intervals I		1,2,7 7 4,5,7
5: 26th Sept (M,W,F)	Regression Review Hurricane Ian Hurricane Ian		
6: 3rd Oct (M,W,F)	Confidence Intervals II Ridge/Lasso Regression Regression Model Comparison	HW1 due HW2 released	4,5,7 4,5,6,12,15 7, 13
Unit III: Regression Model Selection			
7: 10th Oct (M,W,F)	t-statistic, p-value, F-statistic and R^2 AIC, BIC, CV and GCV K-fold CV		7, 13 7, 12, 13 7, 11,12,13
8: 17th Oct (M,W)	Regression Model Selection Overview Test 1: review	Project details due HW2 due	7,11,12,13
Unit IV: Classification Inference			
9: 24th Oct (M,W,F)	Test 1 Introduction Logistic regression	HW3 released	 8,9 8,9
10: 31st Oct (M,W,F)	Regularized classification Probabilistic classification I Probabilistic classification II		12,13 10,11,12 10,11,12
Unit V: Classification Model Selection			
11: 7th Nov (M,W)	Metrics of performance Threshold metrics	HW3 due/ HW4 released	3, 10 3, 10
12: 14th Nov (M,W,F)	Ranking metrics Probability metrics Choosing metric of performance		3, 10 3, 10 3, 10
13: 21th Nov (M)	Test 2 review	HW4 due	
Thanksgiving break			
Course Conclusion			
14: 28th Nov (M,W,F)	Test 2 Project presentation I Project presentation II		 16 16
15: 5th Dec (M,W)	Project presentation III Project presentation IV	Project due	16 16

Learning outcome: After successful completion of this course, you will acquire knowledge to:

1. Understand the basics of statistical learning and its relation to machine learning.

2. Understand the basics of python and using it as a tool to solve problems in statistical learning.

3. Assess the quality of various statistical learning approaches based on various available metrics

4. Understand different problems in supervised learning.

5. Delve deeper into various aspects of linear regression.

6. Understand and implement multiple linear regression

7. Better understand various issues one might face while using linear regression as a tool to understand different properties of data.

8. Understand basics of classification and its relation to regression.

9. Use logistic regression as a tool to solve classification problems.

10. Compare different classification models for your own problem.

11. Use various resampling approaches to make an intelligent choice of a model for your own data science related problem.

12. Use regularization as a way to produce better models.

13. Understand various ways of controlling the complexity of your statistical learning model.

14. Use dimensionality reduction as a tool to simplify your model to achieve better generalization.

15. Interpret data and models in higher dimensions.

16. Apply the concepts learnt in class to problems of practical importance.