STAT-806 Statistical Learning

Credit Hours: 3-0

Prerequisite: None

Aims and Objectives: Statistics is supporting tool, which can assist MS/PhD Mathematics students working on diverse data structures. Moreover, statistical learning can help MS/PhD Mathematics students in modeling the real life applications. This course will help to understand and model the statistical data of diverse structures. This course aims at introducing students to the concepts of statistical learning with focus on regression, classification and clustering. On successful completion of this course, students will know statistical learning, regression, classification and clustering.

Course Contents: Modern data analysis, methods where fewer assumptions (such as a linear relation between response and explanatory variables) are made and where instead data determine the relation. Nearest neighbor methods, kernel smoothing and generalized additive models. Statistical classification, classical classification methods, advanced methods based on modern regression methods. Structures in data, Data mining i.e. Learning from data

Textbook: Hastie T., Tibshirani R., and Friedman, J., Introduction to Statistical Learning, Springer (2013).

Reference books:

1. James, W., Hastie, T. and Tibshirani R. An Introduction to Statistical Learning, with Applications in R, Springer (2003)
2. Trevor, H., Robert T. and Martin W. Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press (2015).
3. Hastier, T.J. and Tibshirani R.J. Generalized Additive Models 2nd Edition.

ASSESSMENT SYSTEM

Nature of assessment	Frequency	Weightage (%age)
Quizzes	Minimum 3	10-15
Assignments	-	5-10
Midterm	1	25-35
End Semester Examination	1	40-50
Project(s)	-	10-20

Weekly Breakdown		
Week	***Section***	***Topics***
1	1 (2.1-2.2)	Supervised learning and variable types
2	2 (3.2)	Linear Regression
3	3 (3.3)	Model/Subset Selection
4	4 (3.4)	Shrinkage Methods
5	5 (3.5)	Principal Components and Partial Least Squares Regression
6	7 (4.2)	Classification with Linear Regression of an Indicator Matrix
7	8 (4.3)	Linear Discriminant Analysis
8	9 (4.4)	Logistic Regression
9	**Mid Semester Exam**	
10	10 (4.5, 12.2)	Separating Hyperplanes/The Support Vector Classifier .
11	11 (12.4-12.6)	Various Discriminant Analysis
12	13 (7.2-7.4)	Bias, Variance, Error
13	14 (7.10-7.11)	Cross validation and Bootstrap
14-15	15 (14.3)-16 (14.3)	Cluster Analysis
16-17	-	Revision
18	**End Semester Exam**	