Nirma University
Institute of Technology, School of Technology
M Tech Computer Science and Engineering (Data Science)
Semester – II

L	T	P	C
3	0	2	4

Course Code	3CS42D103
Course Name	Advanced Statistical Learning

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

1. comprehend the fundamentals of various statistical learning methods
2. interpret and critically evaluate the outcomes of statistical analysis
3. implement statistical learning methods

Syllabus: **Teaching Hours**

Unit I **2**

Introduction: What is Statistical Learning? Assessing Model Accuracy

Unit II **7**

Overview of Supervised Learning: Introduction, Simple Approaches to Prediction, Statistical Decision Theory, Local Methods in High Dimensions, Supervised Learning and Function Approximation, Model Selection and the Bias–Variance Tradeoff

Unit III **9**

Linear Methods for Regression and Classification: Linear Regression Models and Least Squares, Subset Selection, Shrinkage Methods, Methods using Derived Input Directories, Computational Considerations, Linear Discriminant Analysis, Logistic Regression, Separating Hyperplanes, Piecewise Polynomials and Splines, Regularization, General Linear Modelling

Unit IV **7**

Kernel Smoothing Methods: One-Dimensional Kernel Smoothers, Selecting the Width of the Kernel, Local Likelihood and other Methods, Kernel Density Estimation and Classification, Radial Basis Functions and Kernels, Mixture Models for Density Estimation and Classification, Computational Considerations

Unit V **5**

Model Assessment, Selection, Inference and Averaging: Bias, Variance and Model Complexity, Cross-Validation, Bootstrap and Bagging, EM Algorithm

Unit VI **8**

Additive Models, Trees, SVM and Nearest Neighbours: Generalized Additive Models, Tree-based Methods, Boosting Methods, Random Forests, SVM and Kernels, k-Nearest-Neighbour Classifiers, Adaptive Nearest-Neighbour Methods

Unit VII **7**
Unsupervised Learning: K-means, Self-Organizing Maps, Non-negative Matrix Factorization, Independent Component Analysis

Self-Study:
The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:
Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

1. Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J., The elements of statistical learning: data mining, inference and prediction, The Mathematical Intelligencer, 27(2), 83-85.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R., An introduction to statistical learning (Vol. 112, p. 18), New York: springer.
3. Berk, R. A., Statistical learning from a regression perspective (Vol. 14), New York: Springer.
4. Urdan, T. C., Statistics in plain English, Routledge.
5. Haslwanter, T., An Introduction to Statistics with Python, Springer International Publishing.

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list