|  |  |  |  |
| --- | --- | --- | --- |
| Modules | Topics | #Lecture Hours | Reference |
| Module 1:  Introduction | What is Data Science: Life Cycle, Data Visualization, Data Cleaning, Prediction and Classification, Supervised and Unsupervised Learning | 4 | [4] Chapter 1, 2  [3] Chapter 1,3 |
| Module 2:  Dimensionality reduction basics | Matrix Algebra, Gradient of function, Principal component analysis, Singular Value Decomposition | 8 |  |
| Module 3:  Statistic Learning Basics | Population, sample, Central limit theorem, Probability distribution, Variance, Variance- Covariance Matrix for multidimensional random variable, t-distribution, unbiased estimator, hypothesis testing, p-value, F-test, significance level, confidence interval | 10 |  |
| Module 4:  Prediction | Simple Linear Regression, Multiple Linear Regression – estimate of coefficients by ordinary least square (OLS), Statistical inference on regression coefficients, python programming on regression | 6 |  |
| Module 5:  Classification- Supervised | Classification Problem and confusion Matrix,  Logistic Regression(binary/multiclass), K-NN method, Support Vector Machine (SVM), Linear discriminant Analysis (LDA), Decision Tree and Random Forest  Python Programming on Supervised Classification | 10 |  |
| Module 6:  Classification-Unsupervised | K-Means, K-Medoid  Python Programming on Unsupervised classification | 2 |  |
|  |  |  |  |

Int

B

**Books**

1. Econometrics: Fifth Edition, Jeffrey M. Wooldridge, Michigan State University
2. The Elements of Statistical Learning Data Mining, Inference, and Prediction, Trevor Hastie Robert Tibshirani Jerome Friedman, Springer, Second Edition
3. Data Science from Scratch, Joel Grus, O’Reilly, Second Edition
4. Doing Data Science, Straight talk from the front line, [Cathy O'Neil](https://www.ebooks.com/en-us/author/cathy-oneil/710773/), [Rachel Schutt](https://www.ebooks.com/en-us/author/rachel-schutt/710772/), O’Reilly