BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
INSTRUCTION DIVISION
FIRST SEMESTER 2019-2020
Course Handout Part II

Date: 01/08/2019

In addition to part-I (General Handout for all courses appended to the time table) this portion gives further specific details regarding the course.

Course No. **: CS F320**
Course Title **: Foundations of Data Science**
Instructor-in-charge **: NAVNEET GOYAL (goel@)**

Catalog Description

Data Science is the study of the generalizable extraction of knowledge from data. Unprecedented advances in digital technology during the second half of the 20th century and the data explosion that ensued in the 21st century is transforming the way we do science, social science, and engineering. Application of data science cut across all verticals. A data scientist requires an integrated skill set spanning mathematics, probability and statistics, optimization, and branches of computer science like databases, machine learning etc.

Text Books:

T1. Foundations of Data Science - Avrim Blum, John Hopcroft, Ravi Kannan, January, 2018

T2. An Introduction to Data Science – Jeffrey Saltz and Jeffrey Stanton, Sage Pubilcations, September 2017

Reference Books:

R1. Christopher M. Bhisop, Pattern Recognition & Machine Learning, Springer, 2006.

LECTURE PLAN

Topic	Topic Details	No. of Lectures	Chapter Reference
Course Overview & Introduction to Data Science	1. Motivation/course objectives 2. Some motivating applications 3. Tyes of Data	2	T1 – Ch. 1 T2 – Ch. 1 Class Notes + web resources
High-dimensional data & Curse of Dimensionality	1. Characteristics of High-dimensional data, 2. Curse of Dimensionality (CoD) problem 3. Dimensionality Reduction Technique – PCA & SVD 4. Tensors	6	T1 – Chs. 2, 3
Big Data & Big Data Analytics	1. Big Data - sources & applications 2. Social Media Data 3. Introduction to Big Data Analytics	2	T2 – Ch. 20

Frequentist vs. Bayesian approach to Probability	1. Frequentist Approach 2. Bayesian Approach 3. Prior to Posterior – Bayes' Theorem 4. MLE vs. MAP	2	Class Notes + https://sites.google.com/site/bayestutorial/
Probability Distributions and Mixture Models	1. Exponential family of distributions (Bernoulli, Beta, Binomial, Dirichlet, Gamma, & Gaussian) 2. Mixture Models – Mixture of Gaussians	2	R1 – Ch.2, Appendix B
Optimization Techniques	1. Unconstrained/Constrained optimization 2. Convex Optimization & Lagrange Multipliers 3. Quadratic Programming 4. Primal/dual 5. Kernels	4	Class Notes
Data Preparation & Modeling	1. Data wrangling techniques 2. Introduction to Data Modeling a. Relational model b. NoSQL models	4	T2 – Chs. 5,6
Machine Learning Basics	**Supervised Learning** 1. Linear Regression Models a. Polynomial regression b. Linear basis function models 2. Classification a. Naïve Bayes' Classifier b. Decision Tree Learning c. Logistic Regression d. Artificial Neural Networks c. Support Vector Machines d. Instance-based Classifiers **Unsupervised Learning: Clustering** 1. K-means 2. Expectation Minimization **Probabilistic Graphical Models (PGM)** 1. Bayesian Belief Networks (BBN) 2. Markov Random Fields (MRF) 3. Hidden Markov Models (HMM) **Anomaly Detection Techniques**	8-10	T1 – Chs. 5,7,9 T2 – Ch. 18 R1 – Chs. 1,3,8,9
Time-series Data & Analytics	1. Importance & Characteristics of time series data 2. Sources of time series data 3. Time Series analytics	4	Class Notes + web resources
Distributed Computing Frameworks	1. MapReduce and its variants 2. Spark	2	Class Notes + web resources
Data Visualization	1. Visualization Foundations 2. Visualization Pipeline 3. Scalar, Vector, & Tensor Visualization 4. Visualization Techniques for Spatial, Geospatial, & Time-series Data	4	T2 – Chs. 12,13

Evaluation Scheme:

Component	Duration	Weightage	Date (Time)
Midsem Test (Closed Book)	90 Mins.	30%	
Assignments (02)	Take Home	20%	-
Lab. Component	60 Mins.	10%	TBA
Comprehensive Exam (partly open)	3 Hours	40%	11/12 (FN)

Labs. on R: No structured lab. sessions, but students will be provided with Lab. sheets on important topics.
Notices: All notices will be displayed on NALANDA only.
Chamber Consultation Hour: M, W 5.45 to 6.30 pm (6121-K, NAB)
Makeup Policy: To be granted only in case of serious illness or emergency.
NC Policy: Students securing 10% or less marks will get an NC grade. Also, students in the [10-15] bracket are also likely to get NC.

Instructor-in-charge
CS F320