4. From Algorithms to Z-Scores: Probabilistic and Statistical Modeling in Computer Science Norm Matloff, University of California, Davis (Creative Common Licence) http://heather.cs.ucdavis.edu/~matloff/132/PLN/probstatbook/ProbStatBook.pdf

MCSC105: DATA MINING [3-0-1]

Course Objectives: In this course, the objective is to introduce the KDD process. The course should enable students to translate real-world problems into predictive and descriptive tasks. The course also covers data cleaning and visualization, supervised and unsupervised mining techniques.

Course Learning Outcomes : At the end of the course, the student will be able to

CO1: distinguish between the process of knowledge discovery and Data Mining.
CO2: play with basic data exploration methods to develop understanding of given data
CO3: identify suitable pre-processing method for give problem.
CO4: describe different data mining tasks and algorithms.
CO5: use programming tools (e.g. Weka/Python/R etc) for solving data mining tasks.
CO6: follow formal notations and understand the mathematical concepts underlying data mining algorithms

Syllabus:

Overview: The process of knowledge discovery in databases, predictive and descriptive data mining techniques, and unsupervised learning techniques.

Data preprocessing : Data cleaning, Data transformation, Data reduction, Discretization

Classification: Supervised learning/mining tasks , Decision trees, Decision rules, Statistical (Bayesian) classification, Instance-based methods (nearest neighbor), Evaluation and Validation methods.

Clustering : Basic issues in clustering, Partitioning methods (k-means, expectation maximization), Hierarchical methods for clustering, Density-based methods, Cluster Validation methods and metrics

Association Rule Mining: Frequent item set, Maximal and Closed itemsets, Apriori property, Apriori algorithm.

Readings:

1. J Zaki Mohammed and Wagner Meira, **Data Mining and Analysis: Fundamental Concepts and Algorithms,** Cambridge University Press, 2014.
2. P. Tan, M. Steinbach and V. Kumar, **Introduction to Data Mining**, Addison Wesley, 2006.
3. Jiawei Han and Micheline Kamber, **Data Mining: Concepts and Techniques** (3nd ed.), Morgan Kaufmann, 2011.
4. Charu C Agrawal, Data Mining: The Textbook, Springer, 2015