**Csi5387: Final Exam (Open Book)**

**Overview of concepts, algorithms and techniques that you are responsible for (in no particular order)**

[Please let me know if I have missed anything important!]

1. **Concepts that you need to be able to describe and explain:**
   1. **General Concepts:**

* Optimal Bayes Learning
* Classification, Regression, Concept Learning, Multi-class learning
* Concept, Instances, Classes, Generalization
* Inductive bias
* Occam’s Razor
* Version Space
* Overfitting
* Pruning
* Bias, Variance, Bias-Variance Dilemma
* Sample Complexity
* PAC Learning
* Curse of Dimensionality
* Outlier
* Parametric, Non-parametric
* Class Imbalance, Class Skew
* Nominal, Continuous Features, Feature Space, Instance Space
* No Free Lunch Theorem
* Class Noise, Attribute Noise
* Smoothing
* Linear classifier, nonlinear classifier
* Concepts used in various algorithms: Voting, Centroid, Kernel, Gaussian Mixture Model, Hidden Units, information gain, KKT conditions, Kernel trick, Slack variable
* Statistical concepts: probability distribution, Normal (Gaussian) distribution, posterior probability, prior probability, central limit theorem, conditional independence, Likelihood, Maximum Likelihood.
* Bayes theorem
  1. **Methods**
* Active Learning
* Online Learning
* Data Stream mining
* Semi-Supervised Learning
* Association Rule Mining
* Unsupervised Learning, Clustering
* Reinforcement Learning
* Feature Selection (Filter/Wrapper), Feature Extraction
* Deep Learning
* Bayesian Learning
* Multi-Label Classification
* Multi-task learning
* Probabilistic Model: Discriminative, Generative
* Ranking Classifiers, Scoring Classifiers
* Lasso
* Big Data Analysis
* Mining Social Networks
* Trust, Provenance and Privacy

1. **Algorithms that you need to understand in detail**

* Decision Trees
* Multiple Layer Perceptrons
* Naïve Bayes
* K-Nearest Neighbours
* Support Vector Machines
* Ensemble methods: Baggin, Boosting, Random Forests, Stacking, Error Correcting Codes
* Clustering algorithms: k-means, single/complete/average link clustering, EM Algorithm

1. Evaluation metrics/methods that you need to understand in detail

* Evaluation Metrics:
  + true positive, true negative, false positive, false negative, true positive rate, false negative rate
  + Accuracy, Error rate
  + Precision, Recall, F-Measure
  + ROC Analysis, Area Under the Curve
* Re-sampling:
  + Cross-validation, Stratified Cross validation, k-fold cross validation
  + Leave-one-out (Jacknife)
  + Bootstrapping
* Statistical Testing:
  + General concepts: hypothesis testing, significance test, confidence interval, p-value, critical value, omnibus test, post-hoc test
  + T-test
  + Sign test
  + McNemar test
  + Wilcoxon Signed Rank test
  + ANOVA
  + Friedman Test
  + Tukey’s test
  + Nemenyi’s test