

Study guide for CS584 midterm exam. This list of topics is by no means comprehensive. It is meant to outline the major topics that we covered so far. In general, you are responsible for anything that we covered in class, homework assignments, as well as the corresponding chapters in the textbook.

Format of the exam: Problem solving, True/False, Multiple-choice, and Analysis questions (HW1 Q1-Q2 are examples of this type of questions).

Chapter 1. Introduction

- Data mining overview – understand what each task does and when to use them.
 - Classification
 - Clustering
 - Association Analysis
 - Anomaly Detection

Chapter 2. Data

- Common issues with data
- Data preprocessing
- Similarity/distance measures

Chapters 4 & 5 (relevant sections covered so far) – Classification

- Basic concepts
- Different classifiers (what they are, how they work, strengths/weaknesses, choice of classifier given specific problems/datasets, etc.)
- You should know the following classifiers in depth, e.g. you should know how to build a classifier from scratch and/or classify a test instance.
 - Decision trees
 - Naïve Bayes
 - Nearest Neighbor
- Model evaluation
 - Evaluation metrics, evaluation methods, model comparison
- Bias and Variance, Model complexity