

Study Guide for Midterm Exam

You may use one sheet of notes (front and back) and a calculator. You will not be asked any questions about WEKA.

- **Data Preparation** – understand each of the major tasks in data preprocessing and how you would apply them to a given dataset.
- **Data processing** – Given a data set partition it properly using either equi-depth or equi-width portioning. Know the equations for each distance metric and be able to apply them to given data. Be able to calculate central tendencies and correlations between attributes.
- **Association Rules** – be able to extract all frequent itemsets from a transaction database and find the strong association rules. Understand the different metrics that can be applied to association rules to determine if they are interesting.
- **Decision Trees**- Know which metric was used for each decision tree algorithm and how you would use them to build a decision tree.
- **Classification** – Know the formulas for each of the metrics used to evaluate a classifier and be able to apply them to a given confusion matrix.
- **Naïve Bayes** – Given a dataset to use as your training data, use a Naïve Bayes classifier to determine which class a new instance is most likely in.