Homework Assignment 3 (**Individual Assignment**): Midterm Prep

Due date: Wednesday, October 14th 2015. Please hand the homework in to the TA in person or email it to the TA.

All questions listed here are from the online book by Mohammed Zaki and Wagner Meira (which you can download from the class webpage). Several of the questions require data preprocessing. We will also go over a sample midterm in class on October 13th (an in-class assignment).

**Clustering Prep:**

Q13.1

Q13.3 (a) (Note: not required to do part (b))

Q14.1 (Note: while we have not covered Rao’s coefficient in class it is defined in the question. Please answer all parts).

Q14.3

Q15.1

**Classification Prep:**

18.1 (Note: Answer the question as it relates to Naïve Bayes)

18.2 (Note: **Instead of answering the question using Naïve Bayes, answer this question using KNN** classifier – with K = 3. Define a distance metric that equally weights all 3 attributes). You will need to use a form of min-max normalization for the third attribute to ensure equal weights.

Q19.4 (Note: Answer the question as it relates to Information Gain and Gini Index)

Q19.5