CAP 5610: Introduction to Machine Learning
Homework 2

HW 2 is due Tuesday October 15, at Midnight. Check the policy on Homeworks.

Chapter 5. Exercises: 1, 2, 4

Chapter 6. Exercises: 2 (you can use Python), 7
The Optdigits dataset can be found at:
`http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits`

Chapter 7. Exercises: 6,7

Chapter 7. Performed k-means and EM for Gaussian mixture in the *wine* and *iris* datasets (you can use Python).

1. The class information should not be used for clustering but you will use it to assign names to clusters.
2. Apply the k-means and EM algorithm to find clusters (three in both datasets)
3. Assing a cluster name based on the most frequenly occuring class label.
4. For both algorithms, find the number of instances that were put in clusters in which they did not belong.

The datasets can be found in:

- `http://archive.ics.uci.edu/ml/datasets/Wine`
- `http://archive.ics.uci.edu/ml/datasets/Iris`

Chapter 9.
- (From Marsland'09) Suppose that the probability of five events are $P(first) = 0.5$, $P(Second) = P(Third) = P(Fourth) = P(Fifth) = 0.125$. Calculate the entropy. Write down what this means.
- (From Stanford CS 121) Build (by hand not using software) a decision tree to predict the activity of your friend on any future Saturday afternoon from the observed values of Weather, Parents, Cash, and Exam. The data can be found in the class schedule page (decisiontree.xls)

Chapter 10. Exercises: 1, 2