

## **Project 2:**

### The Datasets

Contact Lens data

<http://archive.ics.uci.edu/ml/datasets/Lenses>

Iris data

<http://archive.ics.uci.edu/ml/datasets/Iris>

### The Algorithms

1. Clustering
  - a. Simple KMeans
2. PCA
3. SVM
4. Random Forest

Run SVM:

1. Which kernel works better? Why?
2. How did the SVM Compare to the classifiers from Project 1 in terms of training time and performance?

Run PCA and then run SVM on the reduced data:

1. How many principal components did you pick? Why?
2. How did the SVM perform on the reduced data compared to the original data?
3. How much of the variance in the data is described by the first two or three principal components? Show visually.

Run AdaBoost or Random Forest:

1. How did the boosting or bagging compare to the J48 results from Project 1?

Run clustering (k-means) and then apply decision tree and SVM on clustered data.

1. Compare the performance with the previous results.