1) Using PCA to reduce the number of dimensions in Semi-Conductor Data

If there's some flaw in the manufacturing process, we need to know as soon as possible, so that precious time isn't spent processing a flawed product. Some common engineering solutions find failed products, such as test early and test often. But some defects slip through. If machine learning techniques can be used to further reduce errors, it will save the manufacturer a lot of money.

Dataset: http://archive.ics.uci.edu/ml/datasets/SECOM

- Handle the missing values in the data.
- Create the Covariance Matrix for all the features in the data.
- Calculate the Eigen Values and Eigen Vectors
- Plot the amount of variance that is preserved across each principal component and try to figure out the optimal number of principal components that we can get.
- Finally make a list of top 7 principal-components based values of their variances.

2) Face Recognition with PCA

Face recognition is the supervised classification task of identifying a person from an image of his or her face. use a data set called *Our Database of Faces* from AT&T Laboratories, Cambridge available here. The data set contains ten images each of forty people. The images were created under different lighting conditions, and the subjects varied their facial expressions. The images are grayscale and 92 x 112 pixels in dimension.

Even though images are small their feature vector will have 10304 dimensions. Use PCA to represent the same images in a compact way using principal components.

Reshape the matrix of pixel intensities for an image into a vector and create a matrix of these vectors for all the training images. Then apply PCA over that data to attain a lower dimensional representation.

Once you have obtained the reduced data, apply Logistic Regression on the data to this multiclass classification problem. You will be having around 40 class labels in your target variable. Your algorithm should perform one versus all type of classification for every class label.

Follow the cross-validation approach to validate the performance of the model and submit your assignment with average per-class F1 score, recall and precision of the trained classifier.