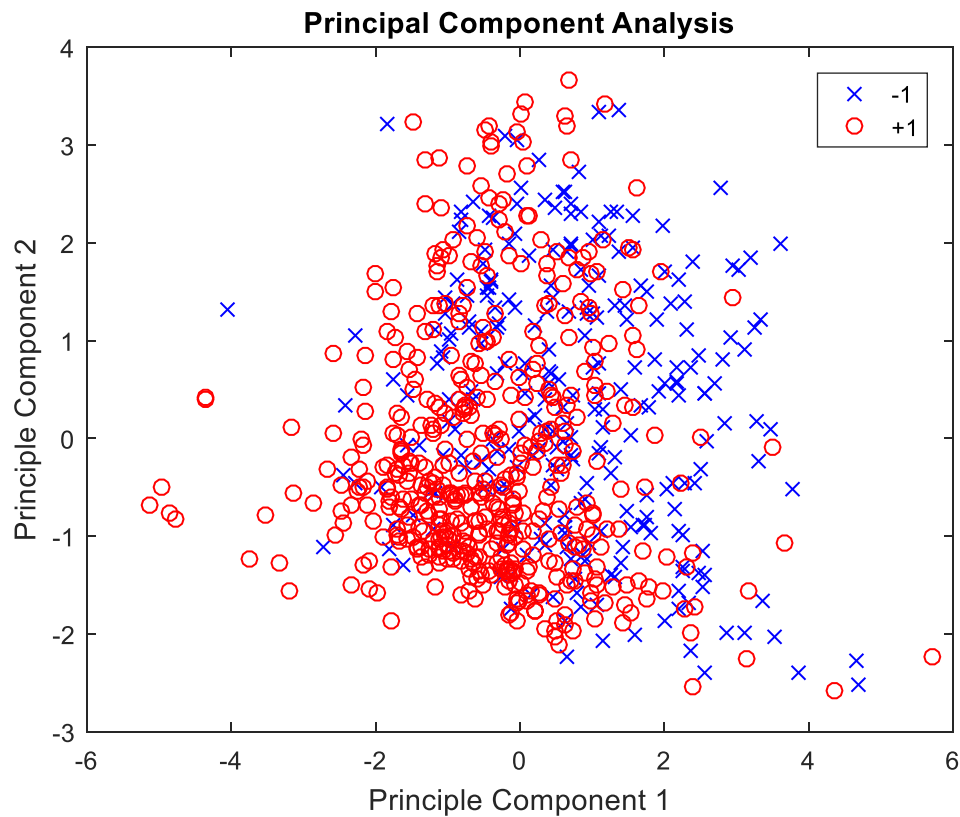
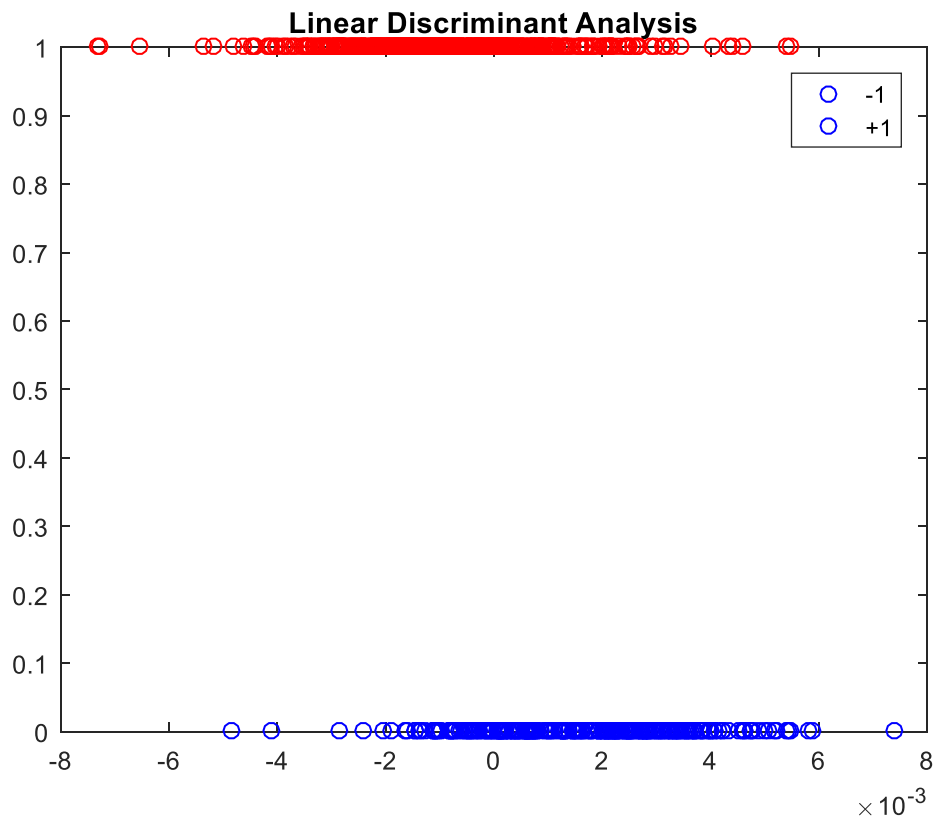


1. PCA



Principal component analysis is an unsupervised method of feature extraction where variance between features is maximized. To perform PCA the data is first standardized by subtracting the data's mean and dividing by its standard deviation. Then the eigenvector of the covariance matrix of the standardized data is computed. The eigenvector with the highest eigenvalue correlates to the feature with the greatest variance. The data is then projected on these features, principal components 1 and 2.

2. LDA



Linear discriminant analysis is a supervised method for dimensionality reduction. It is particularly useful for classification problems. To perform LDA the data is first standardized using the same methods as in PCA. The projection was computed using the following formula:

$$w = S^{-1}(m_1 - m_2)^T$$

where S is the within-class scatter matrix and m_1 and m_2 are the means for each class. The above plot is the result of projecting the data onto the projection matrix w which maximizes the separation between classes.

3. Information Gain

InformationGain	FeatureNumber
0.65092	7
0.34381	6
0.3042	2
0.27709	5
0.14094	8
0.081664	4
0.061825	1
0.05931	3

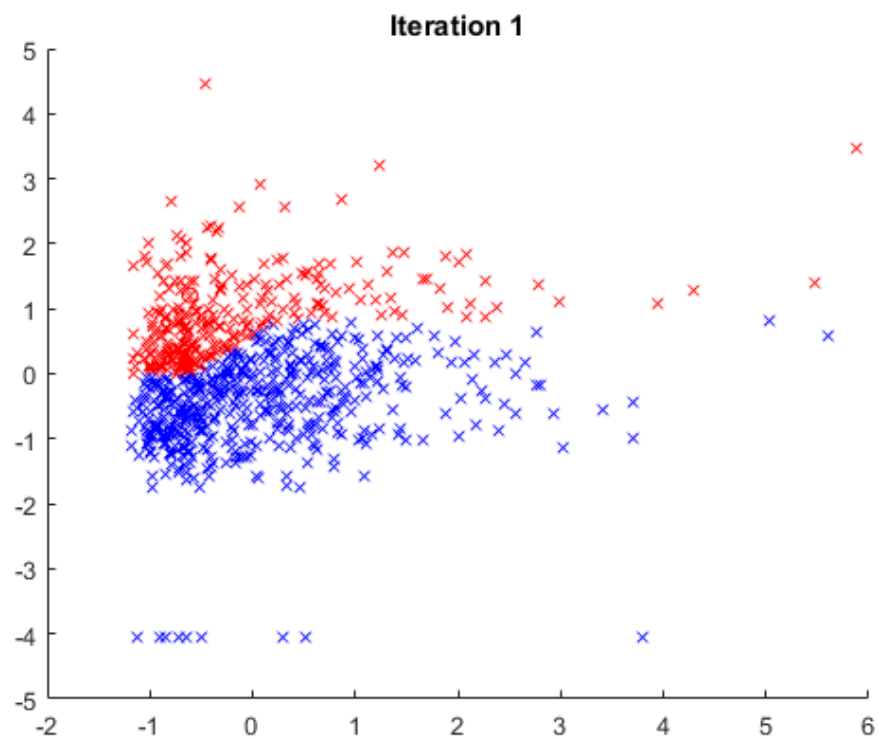
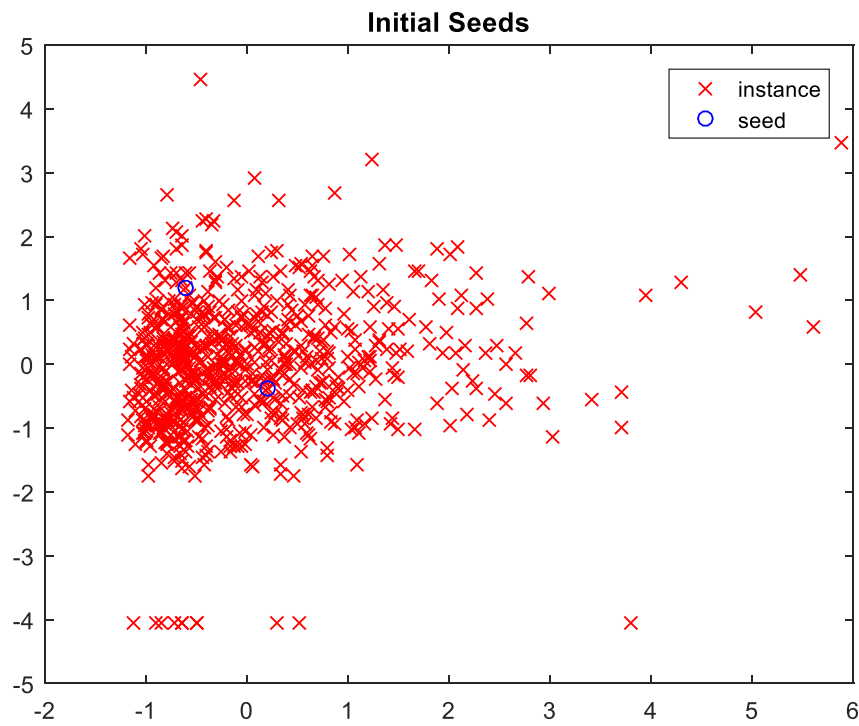
Information gain is method of feature selection that selects features that reduce overall entropy. To perform information gain the data is once again first standardized. Then the weighted average entropy(remainder) of each feature is calculated using the following formula:

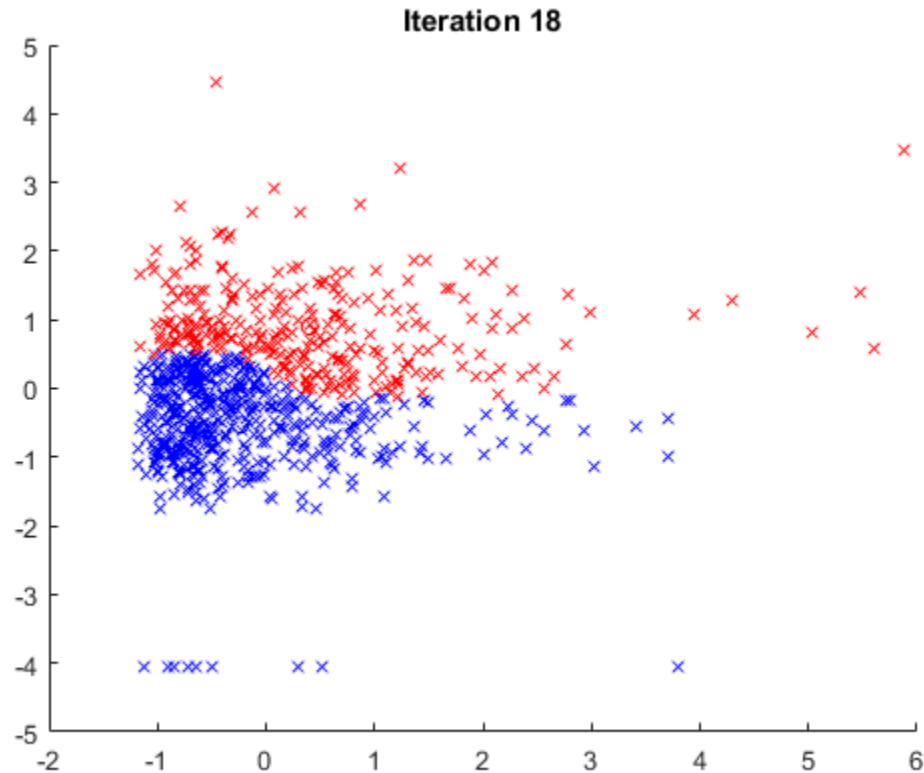
$$remainder(A) = \sum_{i=1}^k \frac{p_i + n_i}{p + n} + H\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

where p_i and n_i are the positive and negative events in feature A respectively. The information gain of each feature is then calculated using the following formula:

$$IG(A) = H\left(\frac{p}{p + n}, \frac{n}{p + n}\right) - remainder(A)$$

4. K-means Clustering





K-means clustering is used for clustering analysis. After standardizing the data and the number, k , of clusters is chosen. K data instances are chosen from the data to act as the initial cluster centers. The first step is to assign each data instance its closest cluster center. The second step is set each cluster center mean of its data instances. These two steps are repeated until convergence. The clustering is terminated when the distance between the old and new cluster centers is less than $2.2204e^{-16}$. This data set converged after 18 iterations.