Assignment 3: PCA, LDA, etc.

1. Choose a binary dataset with at least 100 samples in at least 20 dimensions. Ideally, the data should have continuous valued features. - Mention details of dataset, url, breif description.

2. Determine the principal components. Compute a table and plot in which the x-axis indicates the number of components selected, and the y-axis indicates the variance captured. - Table and graph

3. Project the data onto components capturing at least (a) 75% (b) 85% and (c) 95% variance. - how many components are required for a, b and c.

4. Use any nonlinear classifier on the projected data and plot the test set accuracy versus number of components used. - Table and plot of accuracy on test data vs number of components

5. Run Sanger's M-unit rule for determining the same number of components as in experiment 3. Repeat experiment 4 for projections on these components.
- Table and graph as in point 3 and 4.

6. Determine the LDA components. Choose the same number of components as in experiment 3 above. Use any nonlinear classifier on the projected data and plot the test set accuracy versus number of components used.
- Table and plot of accuracy vs components on test data

7. Determine the Kernel PCA components using a RBF kernel. How many are significant for capturing 80% of the variance ? Plot the projections of the data on the first 3 kernel PCA components.
- Table and 3D plot of projection of data on components.

1. You need to provide a detailed analysis report.
2. Report should include the results observed, graphs, intuitive explainations and conclusion.
3. Submission mode: Moodle.