Module 5: Dimensionality Reduction

Case Study

edureka!

edureka!

Case Study

Objective:

- Understand and practice principal component analysis using scikit learn.

Questions:

1. Scikit learn comes with pre-loaded dataset, load the digits dataset from that collection and write a helper function to plot the image using matplotlib.
   [Hint: Explore datasets module from scikit learn]

2. Make a train -test split with 20% of the data set aside for testing. Fit a logistic regression model and observe the accuracy.

3. Using scikit learn perform a PCA transformation such that the transformed dataset can explain 95% of the variance in the original dataset. Find out the number of components in the projected subspace.
   [Hint: Refer to decomposition module of scikit learn]

4. Transform the dataset and fit a logistic regression and observe the accuracy. Compare it with the previous model and comment on the accuracy.
   [Hint: Project both the train and test samples to the new subspace]

5. Compute the confusion matrix and count the number of instances that has gone wrong. For each of the wrong sample, plot the digit along with predicted and original label.