

Lab Objectives

Implementation of Logistic Regression on PIMAS Diabetes Dataset

Implementation of K means Algorithm

Understanding PCA and Implementing it

PIMAS Diabetes Dataset

Lab Task#01

- Import and Explore PIMAS Diabetes Dataset
- Generate a Bar chart for the count of patients those who have diabetes and those who doesn't have
- Find out those patients who are above 40 years of age and have diabetes.
- Now Split the dataset into training and test data
- Train two logistic Regression Models
 - One without Feature Scaling
 - One with feature scaling
- Find out accuracy of both models
- Perform prediction on any two samples from the test data

K-Means

Lab Task#02

- Load Iris dataset from sklearn
- Train K-Means Algorithm on Iris dataset using the elbow method to find out the best possible number of clusters for this data
- Find out the quality of clustering

Implement the following code to learn how you can visualize clustering with scatter diagram

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.datasets import load iris
# Load the Iris dataset
iris = load iris()
X = iris.data ~
y = iris.target
# Create a KMeans model with 3 clusters
kmeans = KMeans(n_clusters=3, random_state=0).fit(X)
# Get the cluster labels
labels = kmeans.labels_
# Visualize the clusters
plt.scatter(X[:, 0], X[:, 1], c=labels)
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.title('K-Means Clustering of Iris Dataset')
plt.show()
```