Lab - kNN

Q1. Wisconsin Breast Cancer Data – Rescaling

Build a KneighborsClassifier for the WBCD data set and use cross validation to pick the best scaler (unscaled, min-max or z scaler)

- Load the Wisconsin Breast Cancer Data
- Split the data into training and test data.
- Evaluate the KNeighborsClassifier using 5-fold cross validation by finding the cross-validation accuracy for
    1. unscaled data
    2. min-max (MinMaxScaler) scaled data
    3. z-scaled (StandardScaler) data
- Choose the best option.
- Build a KneighborsClassifier using the training data and appropriate scaler.
- Find the test accuracy and confusion matrix.

Q2. WBCD – Finding k

Build a KneighborsClassifier for the WBCD data set and use cross validation to pick the optimal value of k

- Load the Wisconsin Breast Cancer Data
- Split the data into training and test data.
- Scale the data with the appropriate scaler (Q1)
- Find training and validation accuracy for k in range(1,15)
- Plot both on a graph
- Choose the best value of k
- Build a KneighborsClassifier using this value of k
- Find the test accuracy and confusion matrix.

Q3, Q4

Repeat for the iris data.