Assignment2 k-Nearest Neighbors

Due Midnight Apr 16, 2024

1. Construct the kNN classifier to predict Breast Cancer (class 2 = benign, 4 = malignant)
2. Split data into 80% training and 20% testing set (random_state = 1234 for reproducibility)
3. Data Preprocessing:
    a. Convert the classes to a 0 (benign) and 1 (malignant) indicator for using in the classifier
    b. Fill in Missing values, if exist (try using Mode value)
    c. Drop non-value added variables
    d. Standardization→sklearn.preprocessing.StandardScaler
4. Construct KNeighborsClassifier from sklearn.neighbors
5. Use Euclidean distance (Minkowski with 2-norm)
6. Tuning of k value using GridSearchCV (use default 5-fold cross validation), trying k = {1,2,3,4,5,6,7,8,9,10}. Plot graph and report the value of k achieving the highest accuracy.
7. Write the Analysis and Classification report. Submit the report.pdf file including the link to your **colab notebook**.