k-NN Assignment

Code for loading dataset into 2D python list: here

Dataset Preparation(X,y):

Randomly Split the dataset into Training (70%), Validation (10%) and Test (20%)

```
set X_train=[], X_val=[], X_test=[], y_train=[], y_val=[], y_test=[]
//Write code for shuffles your dataset list
1. for each sample a,b in the zip(X,y):
2. generate a random number R in the range of [0,1]
3. if R>=0 and R<=0.7
4. append a in X_train and b in y_train
5. elif R>0.7 and R<=0.8
6. append a in X_val and b in y_val
7. else:
8. append a in X_test and b in y_test
```

KNN Classification:

Use credit card fraud detection data here,

```
K = 5
1. for each sample V in the VALIDATION set:
2. for each sample T in the TRAINING set:
3. Find Euclidean distance between V and T

4. Store T and the distance in list L
5. Sort L in ascending order of distance
6. Take the first K samples
7. Take the majority class from the K samples (this is the detected class for sample V)
8. Now, check if this class is correct or not
9. Calculate validation_accuracy = (correct VALIDATION samples)/(total VALIDATION samples) * 100
```

Note

- Calculate validation accuracy in a similar way for K = 1, 3, 5, 10, 15
- Make a table with 2 columns: K and Validation Accuracy
- Now, take the K with **highest** Validation Accuracy
- Use this best K to determine **Test Accuracy** (Simply replace the VALIDATION set with TEST set)

KNN Regression:

Use weather data here

K = 5, Error = 0
1.for each sample V in the VALIDATION set:
2. for each sample T in the TRAINING set:
3. Find Euclidean distance between V and T
4. Store T and the distance in list L
5. Sort L in ascending order
6. Take the first K samples
7. Take the average output of the K samples (this is the determined output for sample V)
8. Error = Error + (V true output - V determined output)^**2**
9.Calculate Mean_Squared_Error = Error/(total number of samples in VALIDATION set)

Note

- Calculate Mean_Squared_Error in a similar way for K = 1, 3, 5, 10, 15
- Make a table with 2 columns: K and **Mean_Squared_Error**
- Now, take the K with **minimum** Mean_Squared_Error
- Use this best K to determine **Mean_Squared_Error for the Test set** (Simply replace the VALIDATION set with TEST set)

Instruction

● Submit the .ipynb file.
● DO NOT USE LIBRARIES SUCH AS: "Sklearn", "Scikit learning" or for this assignment
● Copying will result in -100% penalty