CSE 475 Lab Assignment – 16/17-05-2022

In this exercise you will

- Use different classifiers
- Use different metrics to analyze the resultant classifications
- Observe and compare the decision boundaries provided by the classifiers
- Tune the hyper-parameters of the classifiers
- Use dimension reduction of features

We will use the following classifiers:

- K Nearest Neighbor Classifier (K)
  https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

- Single and Multilayer Neural Network (#Iterations, #layers, #Neurons, activation functions, loss functions, Optimizer)
  https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

- Support Vector Machine (Kernels)
  https://scikit-learn.org/stable/modules/svm.html
  https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

- Logistic Regression
  https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

- Decision Tree and Random Forests
  https://scikit-learn.org/stable/modules/tree.html
  https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html
  https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

- Naïve Bayes Classifier
  https://scikit-learn.org/stable/modules/naive_bayes.html
  https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

Data

Iris Dataset
https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html

Tasks

1. **Load the dataset**; choose **two** suitable features and plot a 2D image – using 3 different colors for the species.

2. **Classify the data**; for each classifiers given above:
    a. Initialize the classifier with your choice of hyper-parameters
    b. Fit the data
    c. Predict the class of test data
    d. Assess the prediction with confusion matrix
    https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
3. **Observe the decision boundary**
    a. Generate a test data set with all the points in a 2D grid:

```
x = np.linspace(0, 7, 70)
y = np.linspace(0, 3, 30)

#generate all possible value pairs of x and y
xx, yy = np.meshgrid(x, y)
length, width = xx.shape
n_test = length*width
#make a single column of feature values
f1 = xx.reshape(n_test, 1)
f2 = yy.reshape(n_test, 1)
#make test data with above two features
test_data = np.concatenate([f1, f2], axis = 1)
print(test_data.shape)
```

    b. Predict the class of each data point using a classifier
    c. Plot each class with a distinguished color and describe the decision boundary

4. **Tune the hyper-parameters** for each classifiers and report the optimal classification for each classifier.
5. **Perform dimension reduction** on Iris dataset using **PCA**
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html
    a. Apply PCA
    b. Take one or two principal components
    c. Plot the data
    d. Apply the classifiers to the reduced dataset and observe the results