

Programming Assignment 2: Classification Task and Performance Evaluation (10 points)

Shen-Shyang Ho (Dr.)

October 2, 2022

- In this assignment, you will be using the dataset assigned to you in Assignment 1.
 - You will be assigned three classification methods from the following classification methods: **Naive Bayes Classifier, Support Vector Machine (SVM), Decision Tree, Neural Network, Random Forest, Adaboost**
 - Scikit-learn (https://scikit-learn.org/stable/user_guide.html) will be used in this assignment
1. Convert all the images in the dataset to grayscale pixel intensity histograms. (These will be the vector representations of the images).
 2. Split dataset into a training set and a test set. For each class, perform a training/test split of 80/20.
 3. (Model Selection) Perform 5-fold cross-validation on the training set for k-Nearest Neighbor Classifiers such that $k = 1, 3, 5, 7$ on the dataset. (2 points)
 - Plot a graph (x-axis: k ; y-axis: validation accuracy (%)). Which k has the highest accuracy? (1 points)
 - Use the k value with the highest accuracy for your k-Nearest Neighbor classifier. What is the test accuracy? (1 point)
 4. (Performance Comparison) Perform 5-fold cross-validation on the 4-class classification (ignore negative class) using the three assigned classification methods (if you are assigned SVM - use Gaussian kernel and $C = 10$). For neural network (MLPClassifier), you will use default parameters except for learning_rate, you will use 'adaptive'. Plot the confusion matrices for the three approaches (clearly label the classes) using the test set (If you use code from any website, please do proper referencing. You will get 0 point for this assignment without proper referencing) (3 points)
 - Based on the confusion matrices (on the test set), which do you think is the best method? Why? (1 point)
 - Based on the validation accuracies (from the 5-fold cross-validation) for the three methods. Which is the best method? (0.5 point)
 - Compute the test accuracies for the three methods. Which is the best method? (0.5 point)
 - Compute the F-measure for the three methods on the test set. Which is the best method? (1 point)