CAP4103 / CAP6101 Guided Learning #2: Machine Learning Algorithms

This assignment allows you to practice the implementation of classic machine learning algorithms using a simple face recognition system. You've been provided with several files to get you up and running (refer to lecture for a description of these files):

- main.py
- get_images.py
- get_landmarks.py
- shape_predictor_5_face_landmarks.dat
- shape_predictor_68_face_landmarks.dat

You will first need to install the following Python packages: OpenCV (cv2) and Dlib. Then, experiment with ten different sets of classifier/parameter combinations (e.g., [kNN, k = 5], [kNN, k = 10], [NB], [SVM, kernel=linear], etc.). You may also make additional changes, like changing the number of face landmarks or the computation of the features. Your goal is to improve the performance of the system with each subsequent experiment. For example, first, you may decide to run the code as is, and then in the next attempt, change the classifier. Or, you may start with the k-NN classifier, and with each experiment, change the value of k. Only record the changes made in each experiment if performance improves. Record any observations that you feel were important for improving the next experimental attempt.

You will complete these steps for each dataset independently, recording your work using the format below.

    A. **1-10) (Caltech):**

- Classifier:
- Parameters:
- Additional changes:
- Accuracy:
- Observations:

    B. **11-20) (SoF):**

- Classifier:
- Parameters:
- Additional changes:
- Accuracy:
- Observations:

The classifier and its parameters which maximized performance for each dataset are sometimes referred to as models, i.e., a model is a machine learning algorithm trained to identify data patterns. After finding the models that maximize the accuracy of user authentication for the Caltech and SoF datasets (M_c and M_s, respectively), apply M_c to the SoF dataset, and M_s to the Caltech dataset to answer the following questions:

1. Did the classification model M_c generalize to the SoF dataset? If yes, explain why you think it did generalize. If no, explain why you think it did not.

2. Did the classification model M_s generalize to the Caltech dataset? If yes, explain why you think it did generalize. If no, explain why you think it did not.

Use the template GL2-Answers.docx to record your answers.