CS349 Final Project

Option 3: Human Activity Recognition Dataset

Problem: Predict the activity category of a human.

Dataset information:

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

Implementation:

After donwloading the dataset, the first step is to generate training and testing set. Since dataset split train and test set so I only need to import them directly. Next, I write three helper functions: run_model, plot_confusion_mat, model_report. The run_model function can perform different models, it trains the model and does predictions. This function will return a dictionary result contains training time, testing time, predictions, accuracy and trained model. plot_confusion_mat is a function to plot confusion matrix and model_report is a function display the classification report. I tested six models from sklearn: 1. Logistic Regression 2. SGD 3. Kernal SVM 4.Decision Tree 5. MLP 6.KNN. Below is a table for some testing results:

Result/Model	Logistic Regression	SGD	Kernal SVM	Decision Tree	MLP	KNN
Training time	13.694072 s	4.656394 s	26.023740 s	0.169547 s	93.53735 s	0.669222 s
Testing time	0.000268 s	0.000490 s	0.069683 s	0.000338 s	0.002549 s	0.178663 s
Accuracy	0.963013	0.965390	0.965728	0.951812	0.967764	0.964371

From the comparison table, we can see that Decision Tree model has least amount of training time. Logistic Regression model has the least amount of testing time. MLP(Multi-layer Perceptron) has the best accuracy among all six models but it takes very long time to train the models.

For plots of confusion matrix of each model and classification report, please check my jupyter notebook file in my github repo.

Link: https://github.com/dinvincible98/CS349_Final_Project/blob/main/HAR.pdf