Final Project

Data

<https://www.kaggle.com/datasets/uciml/human-activity-recognition-with-smartphones/data>

The Human Activity Recognition database was built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. The objective is to classify activities into one of the six activities performed.

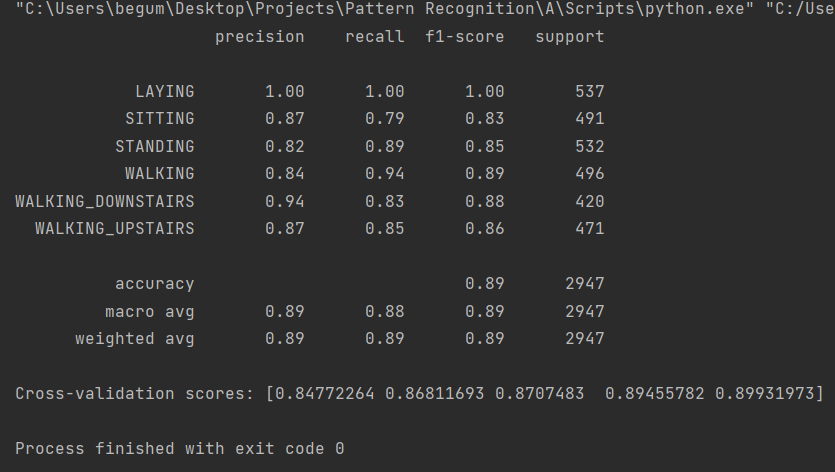
The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer 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.

Codes

I wrote the codes using Python. It uses the Pandas library to read two CSV files, train.csv and test.csv, containing training and testing data, respectively. It separates the features (X\_train and X\_test) and labels (y\_train and y\_test) from the training and testing datasets. It uses the SelectKBest method from scikit-learn with the f\_classif scoring function to select the top 100 features based on their ANOVA F-statistic. It creates a Random Forest Classifier model with 100 trees (n\_estimators=100) and fits the model on the selected features from the training data. It uses the trained model to make predictions on the selected features from the test data and prints a classification report, which includes precision, recall, F1-score, and support for each class. It performs 5-fold cross-validation on the training data using the Random Forest model and prints the cross-validation scores. In summary, this code demonstrates a workflow for a classification task using a Random Forest Classifier with feature selection and cross-validation. The feature selection is done using ANOVA F-statistic to choose the top 100 features. The model is then trained on these features, and its performance is evaluated on a separate test set. Additionally, cross-validation scores are computed to assess the model's generalization performance.

import pandas as pd  
from sklearn.ensemble import RandomForestClassifier  
from sklearn.metrics import classification\_report  
from sklearn.feature\_selection import SelectKBest, f\_classif  
from sklearn.model\_selection import cross\_val\_score  
  
train\_data = pd.read\_csv(r'C:\Users\begum\Desktop\Projects\Pattern Recognition\FinalCSV\train.csv')  
test\_data = pd.read\_csv(r'C:\Users\begum\Desktop\Projects\Pattern Recognition\FinalCSV\test.csv')  
  
X\_train = train\_data.iloc[:, :-1]  
y\_train = train\_data.iloc[:, -1]  
X\_test = test\_data.iloc[:, :-1]  
y\_test = test\_data.iloc[:, -1]  
  
selector = SelectKBest(f\_classif, k=100)  
X\_train\_new = selector.fit\_transform(X\_train, y\_train)  
X\_test\_new = selector.transform(X\_test)  
  
model = RandomForestClassifier(n\_estimators=100, random\_state=42)  
model.fit(X\_train\_new, y\_train)  
  
y\_pred = model.predict(X\_test\_new)  
report = classification\_report(y\_test, y\_pred)  
print(report)  
  
scores = cross\_val\_score(model, X\_train\_new, y\_train, cv=5)  
print('Cross-validation scores:', scores)

Results



The Random Forest Classifier was employed to classify human activities based on sensor data. The dataset was preprocessed, and feature selection was performed using ANOVA F-statistic, retaining the top 100 features. The model was trained on the selected features and evaluated on a separate test set. Additionally, 5-fold cross-validation was employed to assess the model's generalization performance.

Conclusion

The Random Forest Classifier achieved an overall accuracy of 89% on the test set, demonstrating robust performance in classifying various human activities. The classification report provides insights into the precision, recall, and F1-score for each activity class. The model's generalization capability was further confirmed by consistent cross-validation scores, ranging from 84.8% to 89.9% across folds. These results suggest that the Random Forest model, with feature selection, is effective in accurately classifying human activities based on the provided sensor data.