HUMAN ACTIVITY CLASSIFICATION WITH SMARTPHONES PROGRESS REPORT

Section 1

Group 8

Team Members

Mehmet Mert Epsileli - 21502933
Selin Uryan - 21501344
Fırat Ege Akın - 21602166
Burak Korkmaz - 21601296
Ömer Fatih Çelik - 21601387

Introduction and Background Information

The dataset we used contains the recordings of the experiment Human Activity Recognition with Smartphones, which has been carried with 30 participants between 19-48 years old doing their daily activities. The daily activities performed by the participants are WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING. During the recordings, participants wear embedded accelerometer and gyroscope of Samsung Galaxy S II on the waist to capture 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz.The recordings of the participants are separated into 2 groups as 70% for training data and 30% for test data. The training data consists of 7353 rows and 563 columns. The test data consists of 2948 rows and 563 columns. Each row contains a record of a participant for different activities. 561 columns are features of the records, 1 column is the person identifier and the last column is the activity label. Features and labels are separated into two different tables. In order to increase the accuracy and decrease the bias and the variance of the model, we combined train and test data, and split them again by using k-fold cross validation, k is 10 in our case.
Some pre-implemented machine learning projects on Human Activity Classification With Smartphones dataset demonstrates that accuracy of KNN algorithm is 97%, CNN is 99.3% and SVM is 97%. We decided to implement SVM algorithm on the dataset because according to internet search SVM algorithm is a good choice to implement, if there is not enough information about the data [1]. Also the kernel trick in SVM provides finding the most suitable classifier for our complex multi-class dataset.

We used scikitLearn to implement PCA, SVM algorithm and k-fold cross validation on the dataset. SVM algorithms with different types of kernels give different accuracy results.We also used numpy to fasten the calculations and array operations.

What We Have Done So Far

Support vector machine algorithm uses some mathematical functions which are defined as kernel. There are different types of kernels such as radial basis function (rbf), linear, polynomial and sigmoid. We trained several support vector machine models with these different kernels and compared their prediction accuracy. We used the principal component analysis to reduce the dimensions of the data set while retaining %99 of the total variance. The data set includes 561 features, and after the principal component analysis method it was reduced to 157. We used k-fold cross-validation method to calculate the accuracy where k = 10. The support vector machine kernels we have tested are; radial basis function, linear, polynomial and sigmoid.

Following tables display achieved accuracy rates for each step of k-fold cross validation. Each table is a demonstration of a different kernel.

SVM With Kernel = Radial Basis Function (rbf)		
K-FOLD CROSS VALIDATION RESULTS		
Test Rows	Train Rows	Accuracy
[0,1029]	[1030-10298]	0,969
[1030,2059]	[0,1029]+[2060,10298]	0,919
[2060,3089]	[0,2059]+[3090,10298]	0,888
[3090,4119]	[0,3089]+[4120,10298]	0,936
[4120,5149]	[0,4119]+[5150,10298]	0,972
[5150,6179]	[0,5149]+[6180,10298]	0,952
[6180,7209]	[0,6179]+[7210,10298]	0,954
[7210,8239]	[0,7209]+[8240,10298]	0,953
[8240,9269]	[0,8239]+[9270,10298]	0,958
[9270,10298]	[0,9269]	0,958
AVERAGE		0,9459

Figure 1

SVM algorithm with radial basis function (rbf) and 10-fold cross validation was used to obtain Figure 1.

SVM With Kernel = Linear		
K-FOLD CROSS VALIDATION RESULTS		
Test Rows	Train Rows	Accuracy
[0,1029]	[1030-10298]	0,971
[1030,2059]	[0,1029]+[2060,10298]	0,931
[2060,3089]	[0,2059]+[3090,10298]	0,907
[3090,4119]	[0,3089]+[4120,10298]	0,963
[4120,5149]	[0,4119]+[5150,10298]	0,973
[5150,6179]	[0,5149]+[6180,10298]	0,977
[6180,7209]	[0,6179]+[7210,10298]	0,969
[7210,8239]	[0,7209]+[8240,10298]	0,959
[8240,9269]	[0,8239]+[9270,10298]	0,963
[9270,10298]	[0,9269]	0,977
AVERAGE		0,959

Figure 2

SVM algorithm with linear function and 10-fold cross validation was used to obtain Figure 2. Accuracy average of this implementation is better than other implementations. Therefore, in the

future, SVM algorithm with linear function will be used to compare SVM model with other models.

SVM With Kernel = Polynomial		
K-FOLD CROSS VALIDATION RESULTS		
Test Rows	Train Rows	Accuracy
[0,1029]	[1030-10298]	0,905
[1030,2059]	[0,1029]+[2060,10298]	0,873
[2060,3089]	[0,2059]+[3090,10298]	0,821
[3090,4119]	[0,3089]+[4120,10298]	0,849
[4120,5149]	[0,4119]+[5150,10298]	0,903
[5150,6179]	[0,5149]+[6180,10298]	0,9
[6180,7209]	[0,6179]+[7210,10298]	0,877
[7210,8239]	[0,7209]+[8240,10298]	0,881
[8240,9269]	[0,8239]+[9270,10298]	0,883
[9270,10298]	[0,9269]	0,867
AVERAGE		0,8759

Figure 3

SVM algorithm with polynomial function and 10-fold cross validation was used to obtain Figure 3.

SVM With Kernel = Sigmoid		
K-FOLD CROSS VALIDATION RESULTS		
Test Rows	Train Rows	Accuracy
[0,1029]	[1030-10298]	0,936
[1030,2059]	[0,1029]+[2060,10298]	0,853
[2060,3089]	[0,2059]+[3090,10298]	0,856
[3090,4119]	[0,3089]+[4120,10298]	0,895
[4120,5149]	[0,4119]+[5150,10298]	0,891
[5150,6179]	[0,5149]+[6180,10298]	0,93
[6180,7209]	[0,6179]+[7210,10298]	0,941
[7210,8239]	[0,7209]+[8240,10298]	0,913
[8240,9269]	[0,8239]+[9270,10298]	0,9
[9270,10298]	[0,9269]	0,928
AVERAGE		0,9043

Figure 4

SVM algorithm with sigmoid function and 10-fold cross validation was used to obtain Figure 4.

The best accuracy was achieved from linear kernel according to our tests. This indicates that the linear separations between the label regions result the best model.

What Remains to be Done

In the next stages of the project, we are going to use scikitLearn to implement k-fold cross validation and Neural Network with different activation functions such as identity, relu, logistic and tanh on the dataset. After that, the best accuracy rate which is obtained from Neural Network algorithms with different functions will be compared with the best accuracy rate of implemented SVM algorithm to decide which model is best for our dataset. Lastly, according to KAGGLE page of this dataset, a project which uses KNN algorithm obtain 97% accuracy rate [2]. If we have enough time, we will implement a KNN algorithm and compare all three models to find the best model.

Division of Work

Mert: Neural Network Implementation, PCA Implementation, Research and Progress Report, Proposal, Final Report
Ege: Neural Network Implementation, PCA Implementation, Research and Progress Report, Proposal, Final Report
Fatih: Neural Network Implementation, PCA Implementation, Research and Progress Report, Proposal, Final Report
Burak: SVM Implementation, K-fold Cross Validation Implementation, Research and Progress Report, Proposal, Final Report
Selin: SVM Implementation, K-fold Cross Validation Implementation, Research and Progress Report, Proposal, Final Report

REFERENCES

[1] VibeThemes, “204.6.8 SVM : Advantages Disadvantages and Applications,” *Statinfer*. [Online]. Available: https://statinfer.com/204-6-8-svm-advantages-disadvantages-applications/. [Accessed: 21-Nov-2019].

[2] Uci, “Human Activity Recognition with Smartphones,” *Kaggle*, 13-Nov-2019. [Online]. Available: https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones. [Accessed: 21-Nov-2019].