



Q1

I tried two discriminative models and one generative model. I tried six features and decided to use four(min, max, var, mean) and get an accuracy of 100% in 10-Fold cross-validation. Finally, I used a clustering.

(A)Preprocessing

a) Scaler:

Scaler is not applied here because there is only accelerometer's data. But in Q2 scaler is applied

```
36 scaler=preprocessing.MinMaxScaler()
37 DATA=scaler.fit_transform(DATA)#!!!!!!
38
```

b) Detrend

Detrend is applied here and in Q2.

```
49 fAX=signal.detrend(AX)
50 fAY=signal.detrend(AY)
51 fAZ=signal.detrend(AZ)
```

c) Low-pass-filter

Filter is not applied here but is applied in Q2 because it is all known that the highest running frequency is 5Hz. However, it is not known certain frequency about climbing stairs, driving car or dancing.

d) Results & Analysis: Without “detrend” the accuracy for Q1 is higher, however the clustering performance is worse and the performance of Q2 also worse.

(B)Features extraction

a)mean

b)var

c)min

d)max

e)ptp : (unused)

f) cv =mean/std (unused)

```
70 def sliding_win(DATA):
71     frame_size=200
72     step_size=100
73     for counter in range(0,len(DATA),step_size):
74         dataset_frame=DATA[counter:counter+frame_size,:]
75         dataset_frame_mean=np.mean(dataset_frame,axis=0)
76         dataset_frame_var=np.var(dataset_frame,axis=0)
77         dataset_frame_min=np.min(dataset_frame,axis=0)
78         dataset_frame_max=np.max(dataset_frame,axis=0)
79         dataset_frame_ptp=np.ptp(dataset_frame,axis=0)
80         dataset_frame_cv=np.mean(dataset_frame,axis=0)/np.std(dataset_frame,axis=0)
81
82
83     dataset_frame_features=np.hstack((dataset_frame_var,dataset_frame_mean,dataset_frame_min,dataset_frame_max,dataset_frame_ptp,dataset_frame_cv))
84     if counter==0:
85         dataset_feature_vector=dataset_frame_features
86     else:
87         dataset_feature_vector=np.vstack((dataset_feature_vector, dataset_frame_features))
88     return dataset_feature_vector
89
90
```

(C) Classification

a) Generative – **GNB**

b) Discriminative – **SVM & Random forest**

c) Results:

- When I chose 4 features, the accuracy is always 100% (both split test train and cross validation), the accuracy is as following:

```
[...] SVM cross-validation Accuracy: 1.0
[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]
GNB cross-validation Accuracy: 1.0
[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]
RF cross-validation Accuracy: 0.9988888888888889
[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.9888888888888889]
SVM Split train/test Accuracy: 1.0
GNB -Split train/test Accuracy: 1.0
RF -Split train/test Accuracy: 0.9888888888888889
```

- When I chose 6 features, the accuracy is as following:

```
[...] SVM cross-validation Accuracy: 0.8866666666666667
[0.6555555555555556, 0.6111111111111112, 1.0, 0.8888888888888888, 1.0, 1.0, 1.0, 1.0]
GNB cross-validation Accuracy: 0.9088888888888889
[0.7555555555555555, 0.8111111111111111, 1.0, 0.9555555555555556, 0.7222222222222222]
RF cross-validation Accuracy: 0.8800000000000001
[0.8333333333333334, 0.7555555555555555, 0.9888888888888889, 0.8333333333333334, 1.0]
SVM Split train/test Accuracy: 0.9222222222222223
GNB -Split train/test Accuracy: 0.9333333333333333
RF -Split train/test Accuracy: 0.9555555555555556
```

d) Analysis:

- Discriminative model has higher accuracy here.
- It is **not** true that the more features we use, the better the accuracy.

(D) Clustering (Visualization)

