# Homework 5

1. **Table listing the experiments carried out with the following columns.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Size of the Fixed Length Sample** | **Overlap(0-X%)** | **K-value** | **Classifier** | **Accuracy** |
| 32 | 0 | 24 | SVM | 75.78% |
| 20 | 0 | 24 | SVM | 69.57% |
| 16 | 0 | 24 | SVM | 68.32% |

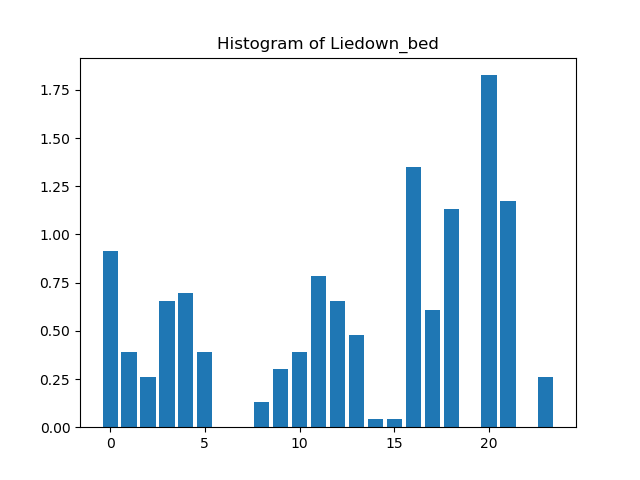
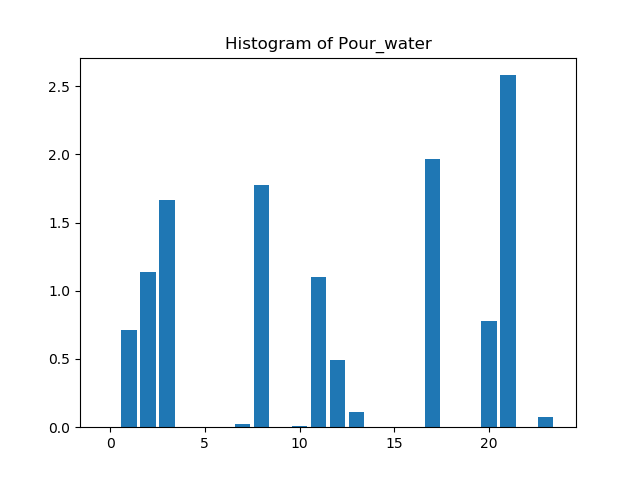
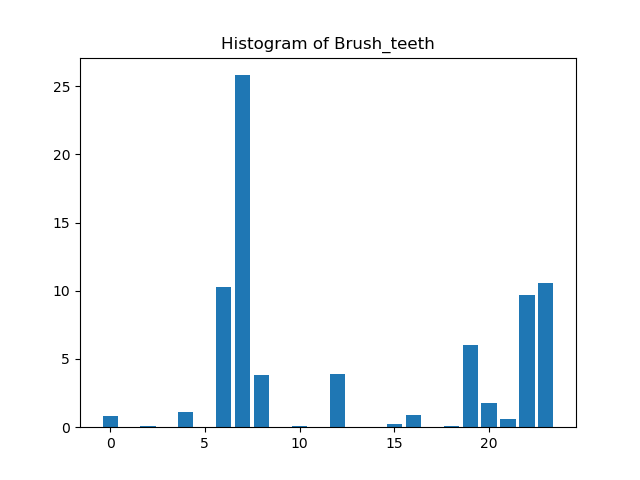
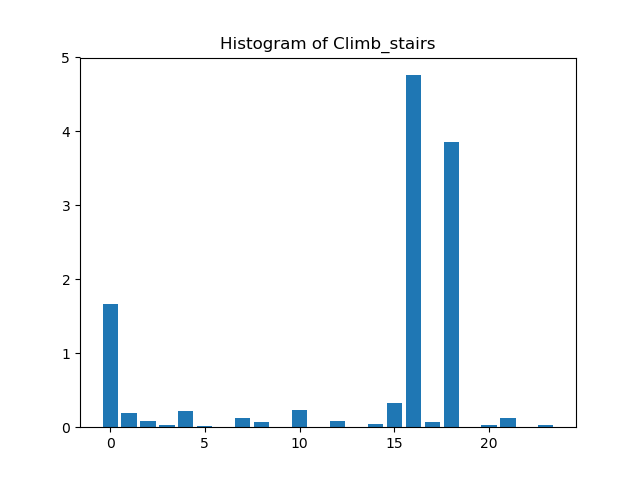
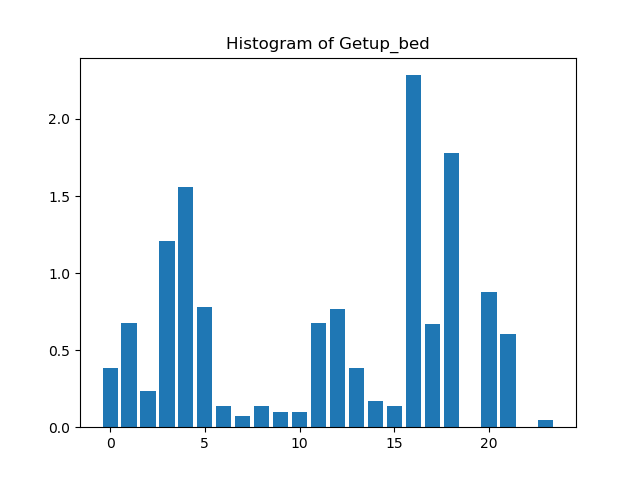
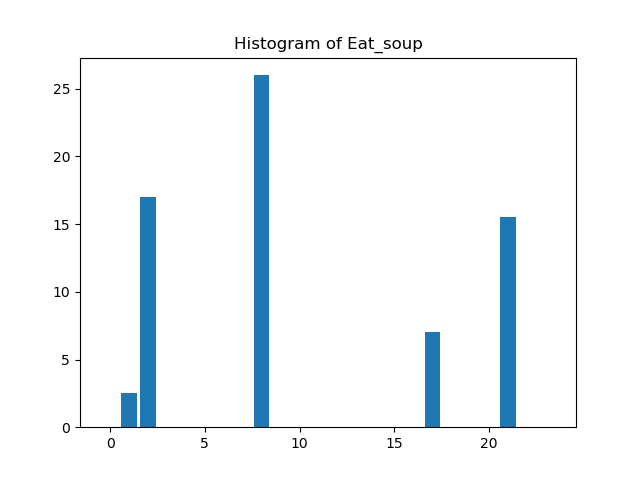
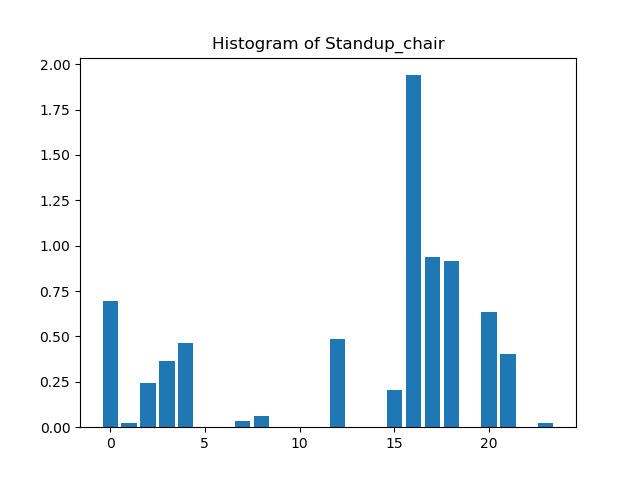
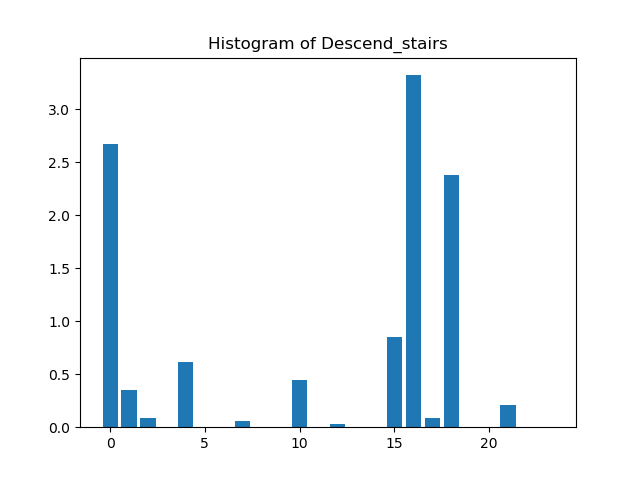
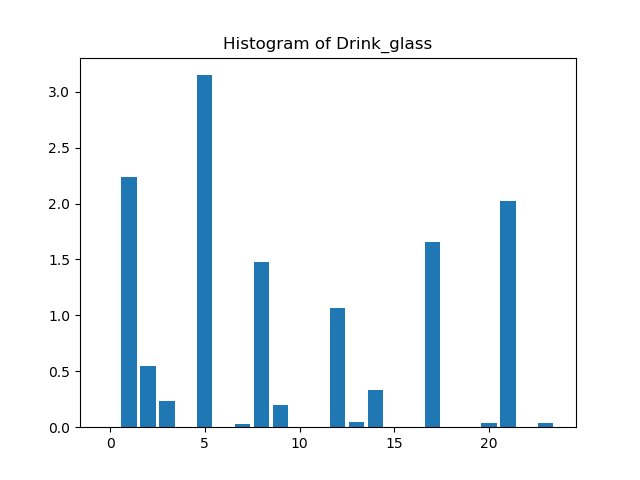
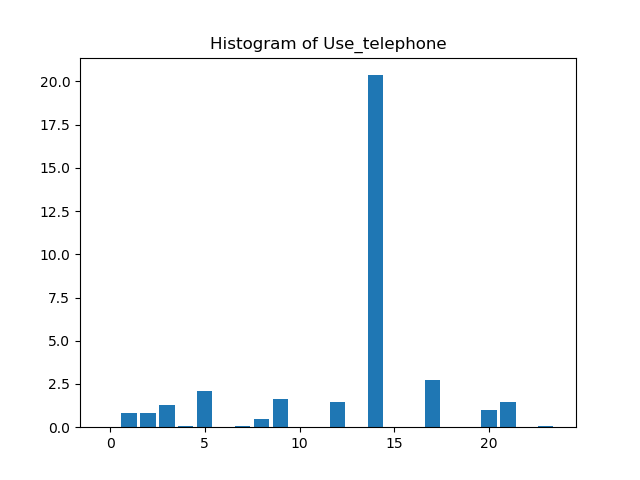
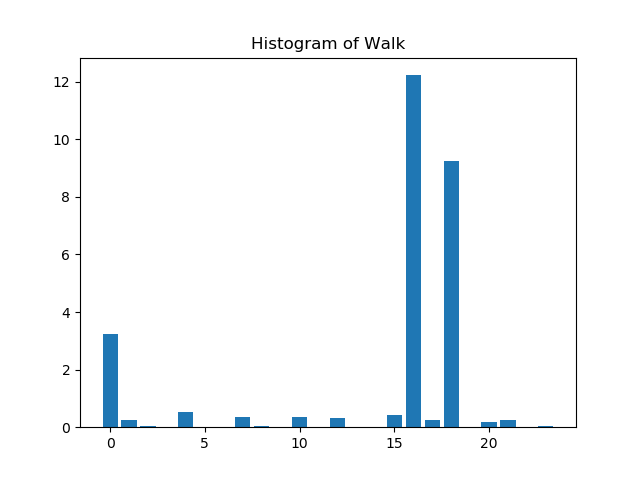
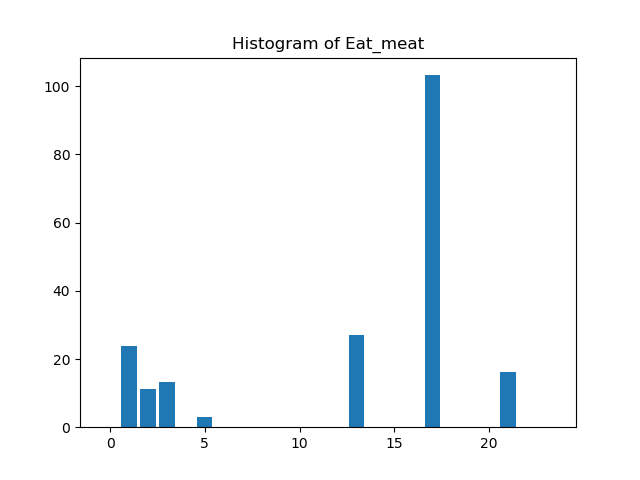
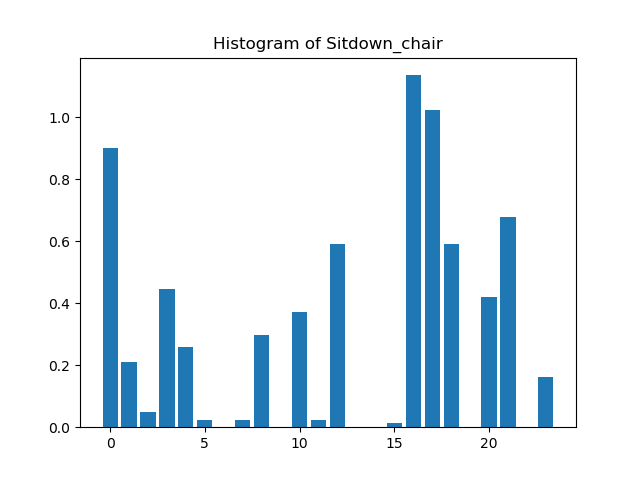
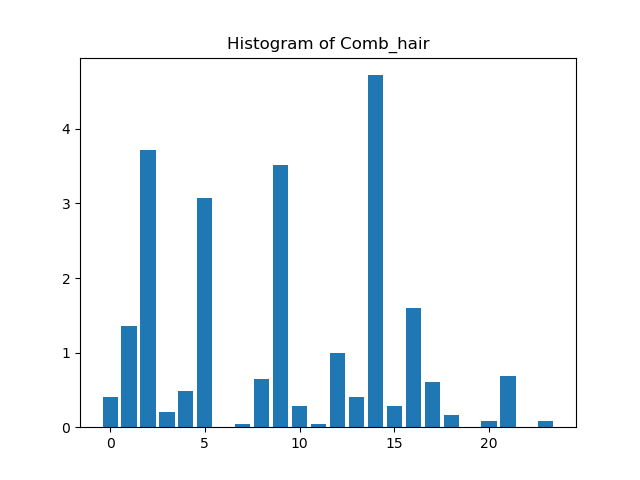
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Size of the Fixed Length Sample** | **Overlap(0-X%)** | **K-value** | **Classifier** | **Accuracy** |
| 32 | 0 | 24 | SVM | 75.78% |
| 32 | 0 | 20 | SVM | 73.29% |
| 32 | 0 | 16 | SVM | 72.05% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Size of the Fixed Length Sample** | **Overlap(0-X%)** | **K-value** | **Classifier** | **Accuracy** |
| 32 | 0 | 24 | Random Forest | 65.22% |
| 20 | 0 | 24 | Random Forest | 64.60% |
| 16 | 0 | 24 | Random Forest | 62.11% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Size of the Fixed Length Sample** | **Overlap(0-X%)** | **K-value** | **Classifier** | **Accuracy** |
| 32 | 0 | 24 | Random Forest | 65.22% |
| 32 | 0 | 20 | Random Forest | 61.49% |
| 32 | 0 | 16 | Random Forest | 66.46% |

1. All K-means use standard K-means.
2. Test-train split percent is 0.8. 80% of files are training set and 20% of files are testing set. We first scan each folder which include files of each activity and there are 14 folders. For each folder(activity), we randomly choose 80% of file as training set and the left 20% as testing files without overlapping. In each file, we first create segments according to the size of the fixed length sample and reshape each segment’s data as [1, size\*3+1]. The last data is the original label of the activity.
3. **Histograms of the mean quantized vector:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Size of the Fixed Length Sample** | **Overlap(0-X%)** | **K-value** | **Classifier** | **Accuracy** |
| 32 | 0 | 24 | SVM | 75.78% |

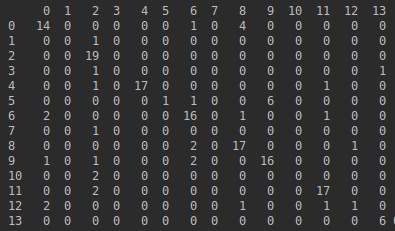
** **

1. **Class confusion matrix from the classifier that you used.**

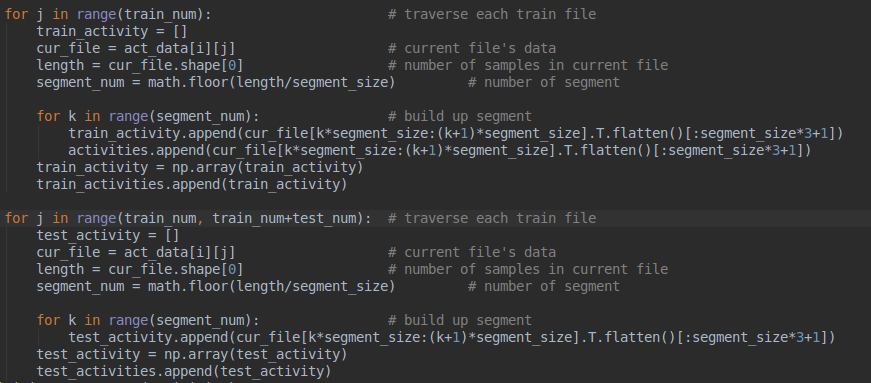
0: Sitdown\_chair, 1: Eat\_meat, 2: Walk, 3: Use\_telephone, 4: Drink\_glass, 5: Descend\_stairs,

6: Standup\_chair, 7: Eat\_soup, 8: Getup\_bed, 9: Climb\_stairs, 10: Brush\_teeth, 11: Pour\_water,

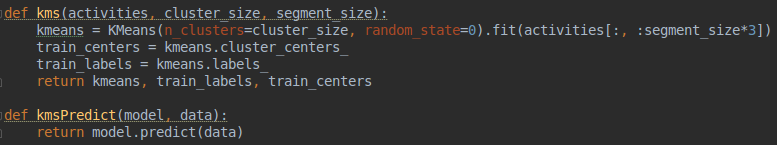
12: Liedown\_bed, 13: Comb\_hair



1. **Code snippets highlighting the following:**
2. **Segmentation of the vector**



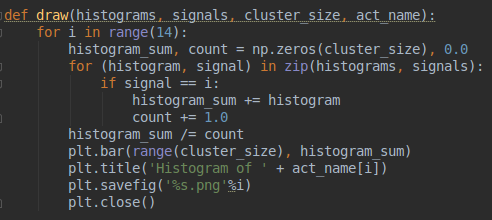
1. **K-means**



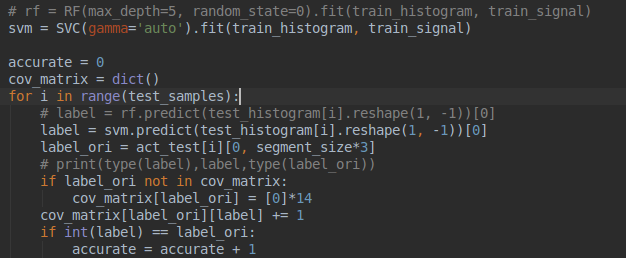




1. **Generating the histogram**



1. **Classification**



1. **Analysis：**

Since each activity’s time/length is different. Thus, we use fixed size of creating multiple segments of each activity. Then, we use k-means to reduce the dimension and create a new feature vector. Though the number of segments at each center is different, the distribution is similar between the same activities and different between different activities. Depending on this, we separate testing data into multiple segments and then find its pattern. Finally, we can calculate the distance between testing pattern and training pattern. We classify the testing activity according to the minimum distance.

According to the confusion matrix, each row means one activity in the original label and each col means the predict label. We can find out the value of the diagonal line is largest which means the correct prediction. Those values which are not on a diagonal line mean incorrect predict the result.

**Code Screenshot:**

