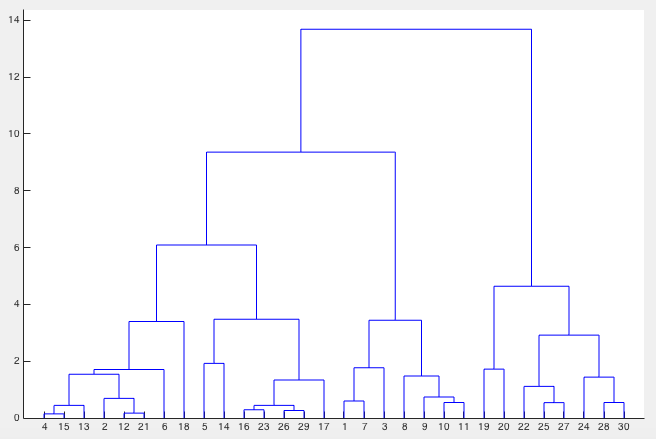
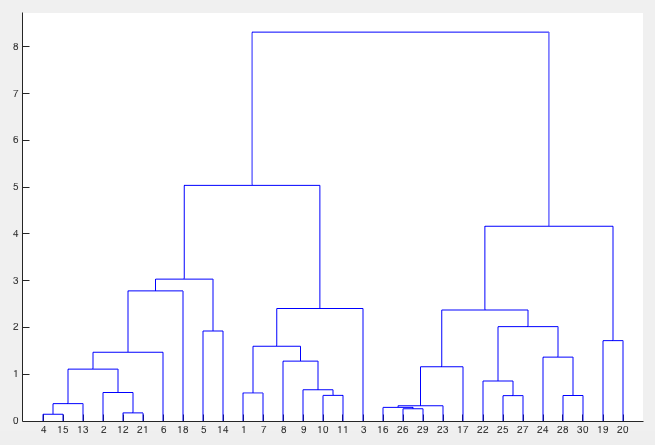
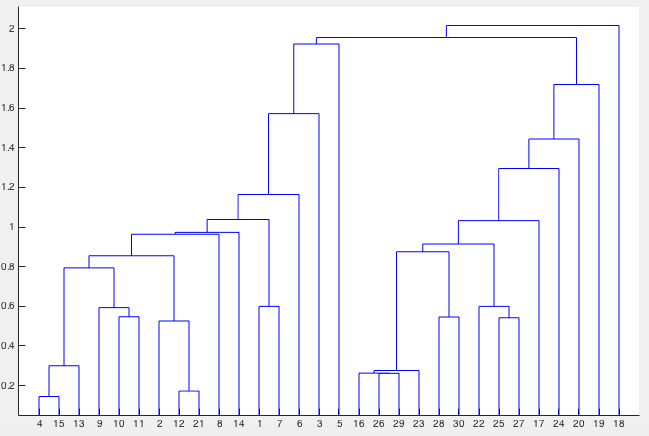
1.

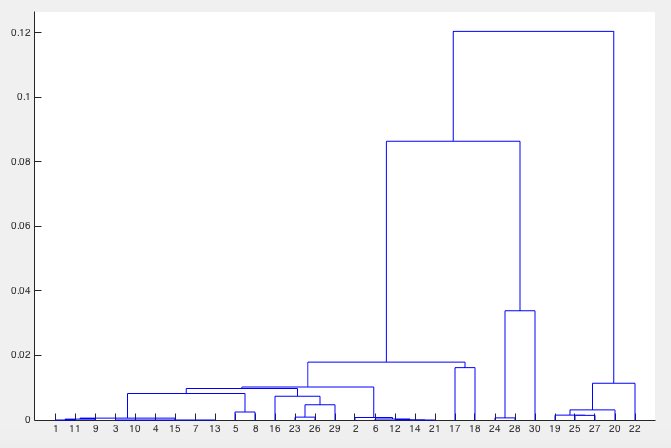
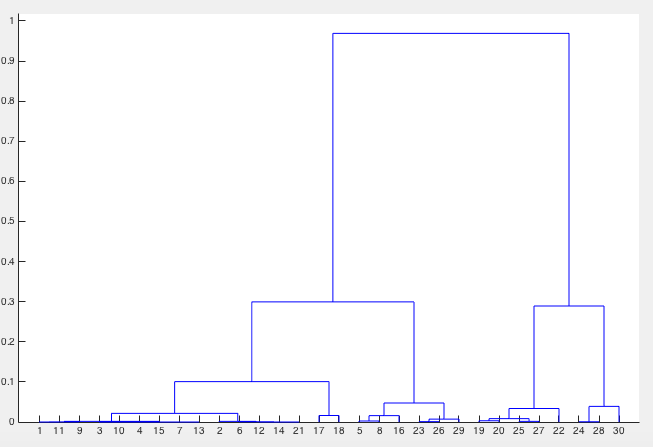
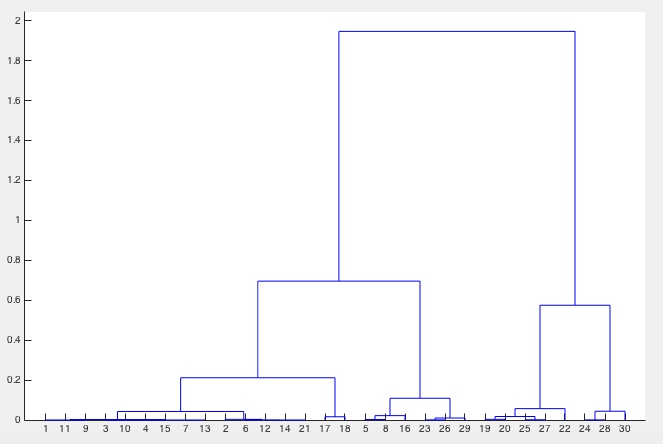
(a) Dendrograms

Euclidean Distance:



Single linkage Complete linkage Average linkage

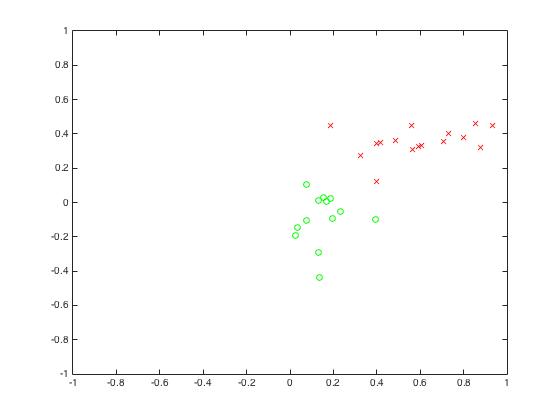
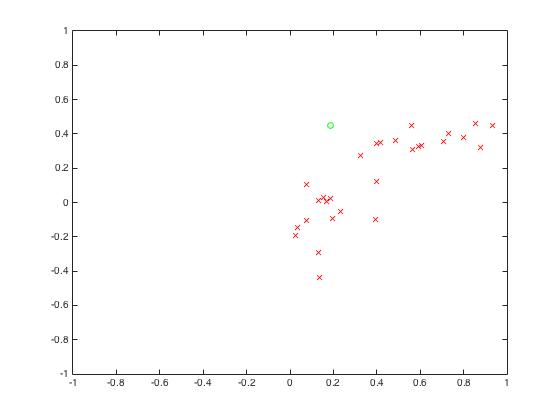
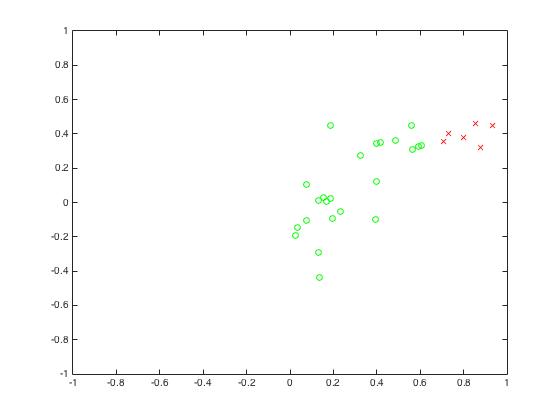
Cosine Similarity:



Single linkage Complete linkage Average linkage

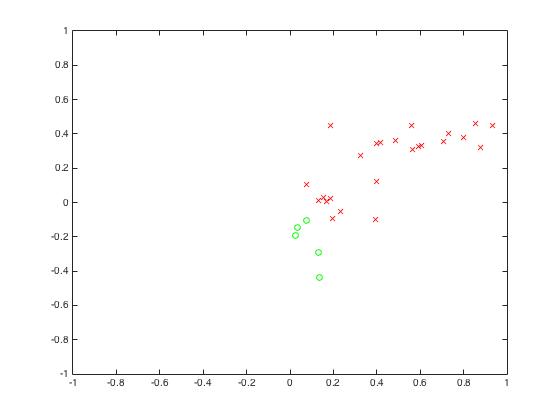
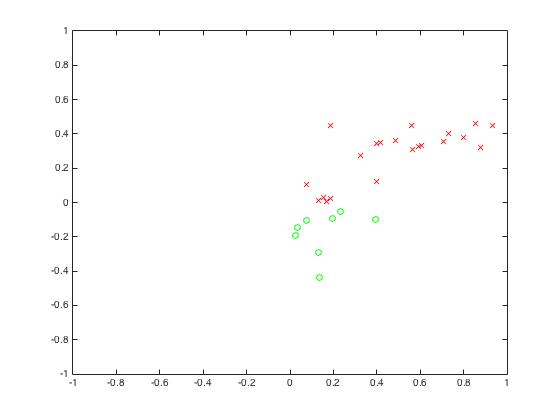
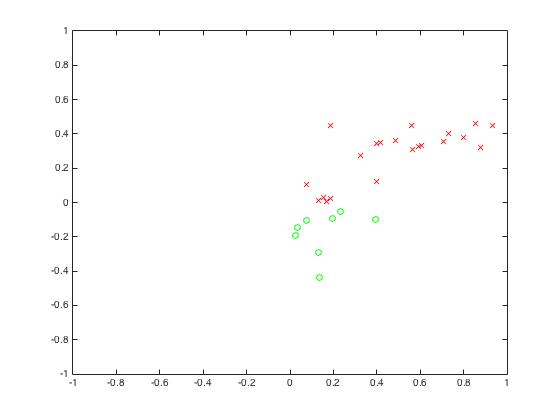
(b)2D cluster

Euclidean Distance:



Single linkage Complete linkage Average linkage

Cosine Similarity:

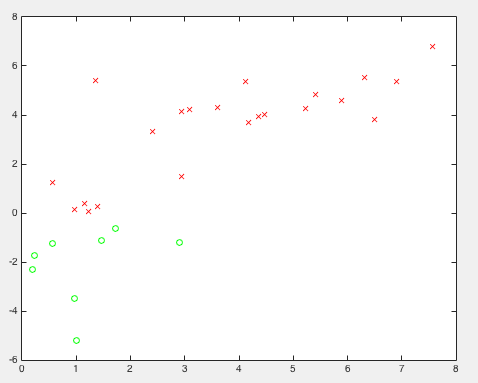
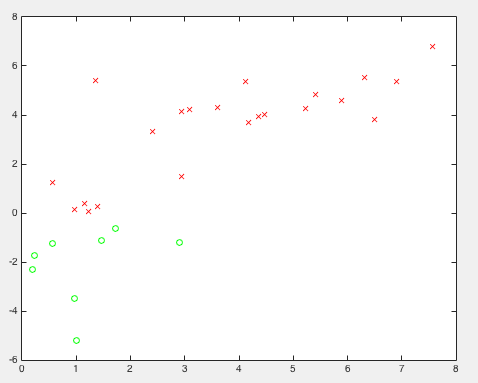


Single linkage Complete linkage Average linkage

(b) Accuracy

|  |  |  |  |
| --- | --- | --- | --- |
|  | Single linkage | Complete linkage | Average linkage |
| Euclidean Distance | 53.33% | 93.33% | 73.33% |
| Cosine Similarity | 66.7% | 76.67% | 76.67% |

(c) Kmeans



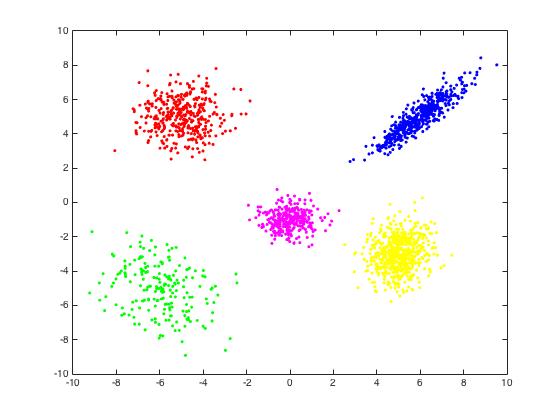
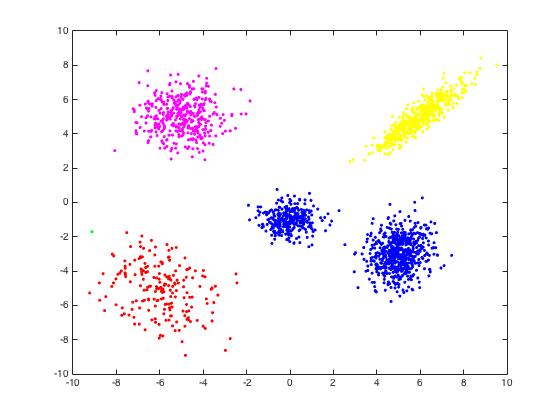
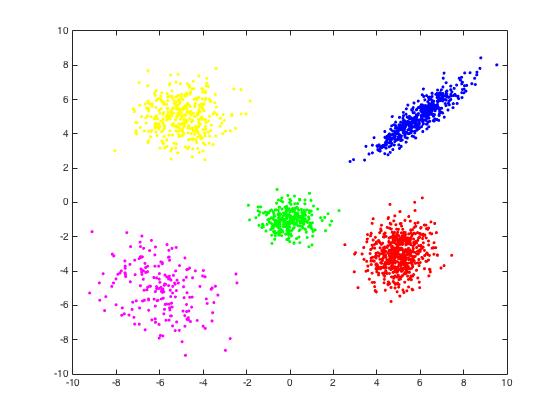
Euclidean Distance Cosine Similarity

(c)Accuracy

|  |  |
| --- | --- |
| Euclidean Distance | 90% |
| Cosine Similarity | 76.67% |

2.

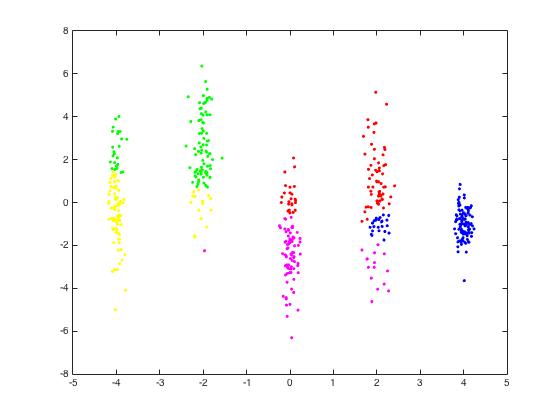
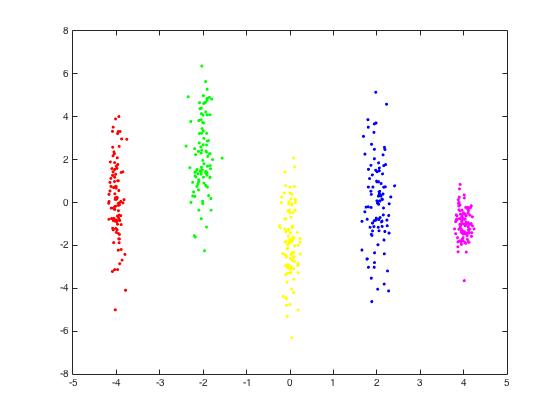
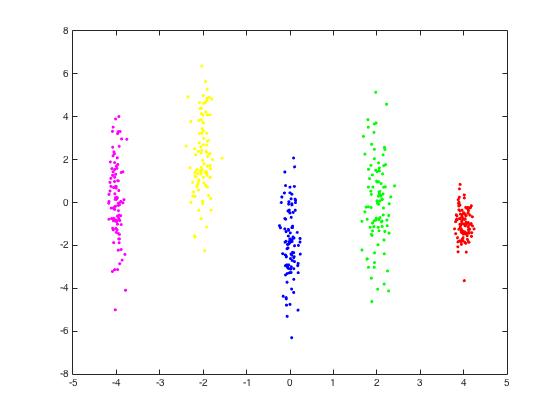
(a)



Kmeans single linkage spectral clustering

Result: spectral clustering = Kmeans > single linkage

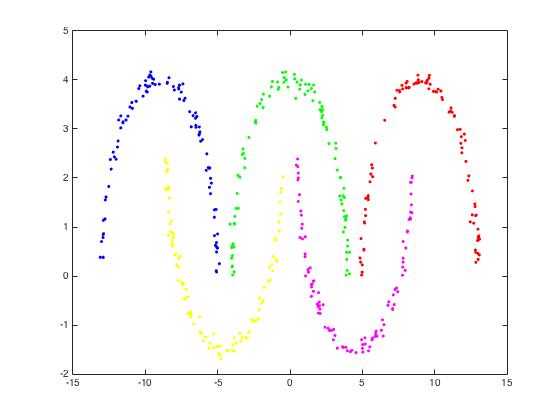
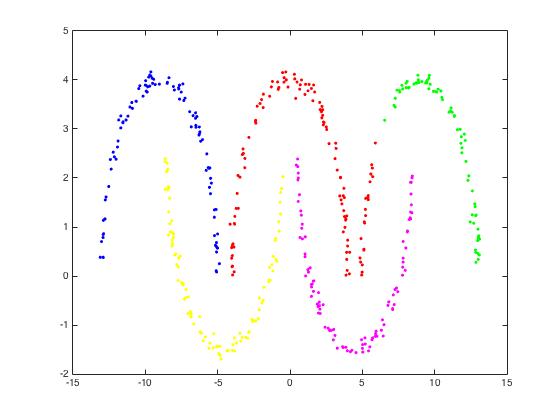
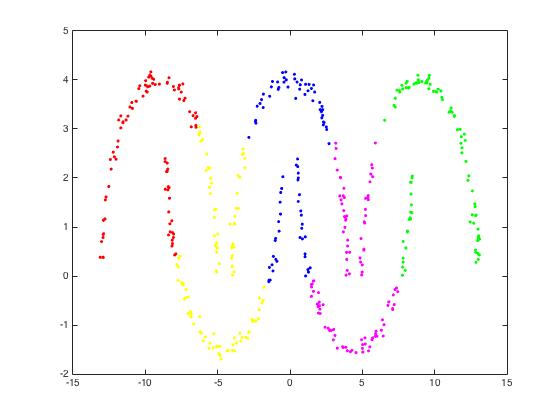
(b)



Kmeans single linkage spectral clustering

Result: Spectral clustering = single linkage > Kmeans

(c)



Kmeans single linkage spectral clustering

Result: spectral clustering > single linkage > Kmeans

In general, spectral clustering is usually better than Kmeans and single linkage.

In the case of (a), it is the most typical case for Kmeans. So it works pretty well like spectral clustering.

In some special case like (b) and (c), the distance within cluster sometimes longer than the distance between different clusters, or sometimes there exists mixture between clusters, then we need spectral clustering to make good classification.

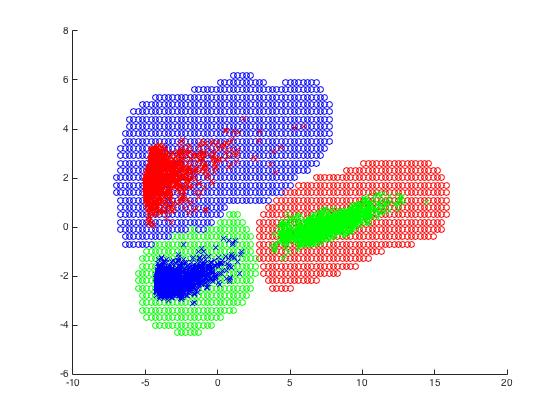
What’s more, if the labels were provided, I think SVM can work well or even better because we can use kernel method which classifies samples in higher dimension space. I think it can solve more complicated classification problems.

3.

(a)PCA (see the code)

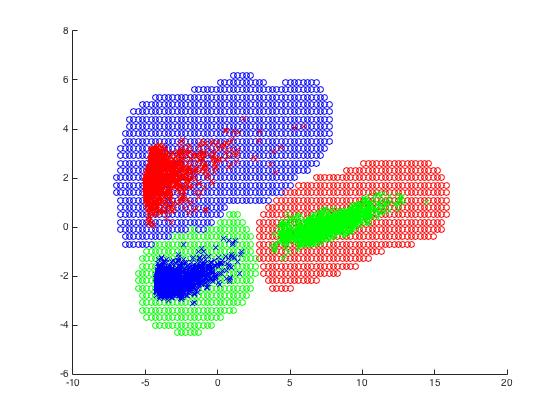
(b)

three decision boundaries:



(c)

three decision boundaries with extra sample



I don’t think an extra sample will result in the change of decision boundary. The reason is that decision boundary is caused by plenty of support vector. It’s not going to cause huge change with just one vector. Also, this new data point may even not a support vector, which couldn’t affect decision boundary at all.

(d)

Test accuracy: 100%

accuracy for each “one versus all training”:

--walking: 100%

--standing: 100%

--laying: 100%

confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
|  | Walking | Standing | Laying |
| Walking | 496 | 0 | 0 |
| Standing | 0 | 532 | 0 |
| Laying | 0 | 0 | 537 |

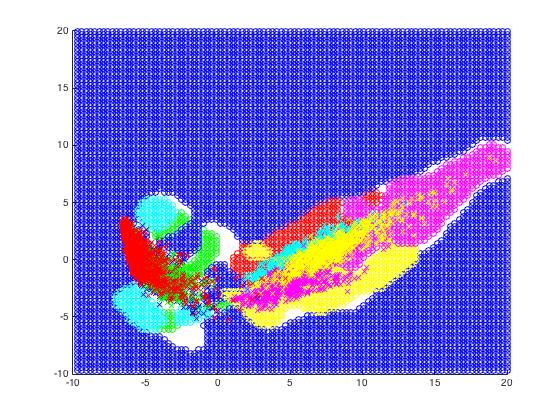
:

(e)

Kmeans

Test accuracy: 99.87% (slightly lower than SVM, but still very good!)

(f)

six decision boundaries

test accuracy: 42.11%

accuracy for each “one versus all training”:

walking: 88.9379%

walking upstairs: 93.5188%

walking downstairs: 88.8316%

sitting: 83.1693%

standing: 82.9318%

laying: 86.359%

confusion matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Upstairs | Downstairs | Sitting | Standing | Laying |
| Walking | 426 | 56 | 200 | 0 | 0 | 0 |
| Upstairs | 2 | 303 | 20 | 0 | 0 | 0 |
| Downstairs | 52 | 21 | 164 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 1 | 2 | 4 |
| Standing | 0 | 0 | 0 | 89 | 168 | 50 |
| Laying | 0 | 1 | 0 | 33 | 10 | 180 |

(g)

the sitting class is the most difficult class to classify.

It makes sense because when people is sitting, it’s really hard to collect correct data.

(h)

the variance is indeed higher when considering all classes. But too many classes makes 2 principal component hard to make different, which cause less accuracy with 6 classes than 3.

(i)

test accuracy: 93.86%

accuracy for each “one versus all training”:

--walking: 99.1856%

--walking upstairs: 97.5908%

--walking downstairs: 98.4502%

--sitting: 97.455%

--standing: 97..7265%

--laying: 100%

confusion matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Walking | Upstairs | Downstairs | Sitting | Standing | Laying |
| Walking | 472 | 5 | 2 | 0 | 0 | 0 |
| Upstairs | 7 | 432 | 20 | 3 | 0 | 0 |
| Downstairs | 10 | 1 | 389 | 0 | 0 | 0 |
| Sitting | 0 | 0 | 0 | 440 | 22 | 0 |
| Standing | 0 | 0 | 0 | 30 | 495 | 0 |
| Laying | 0 | 0 | 0 | 0 | 0 | 537 |