

# ASSIGNMENT 4

## MACHINE LEARNING

Q1a) — — —

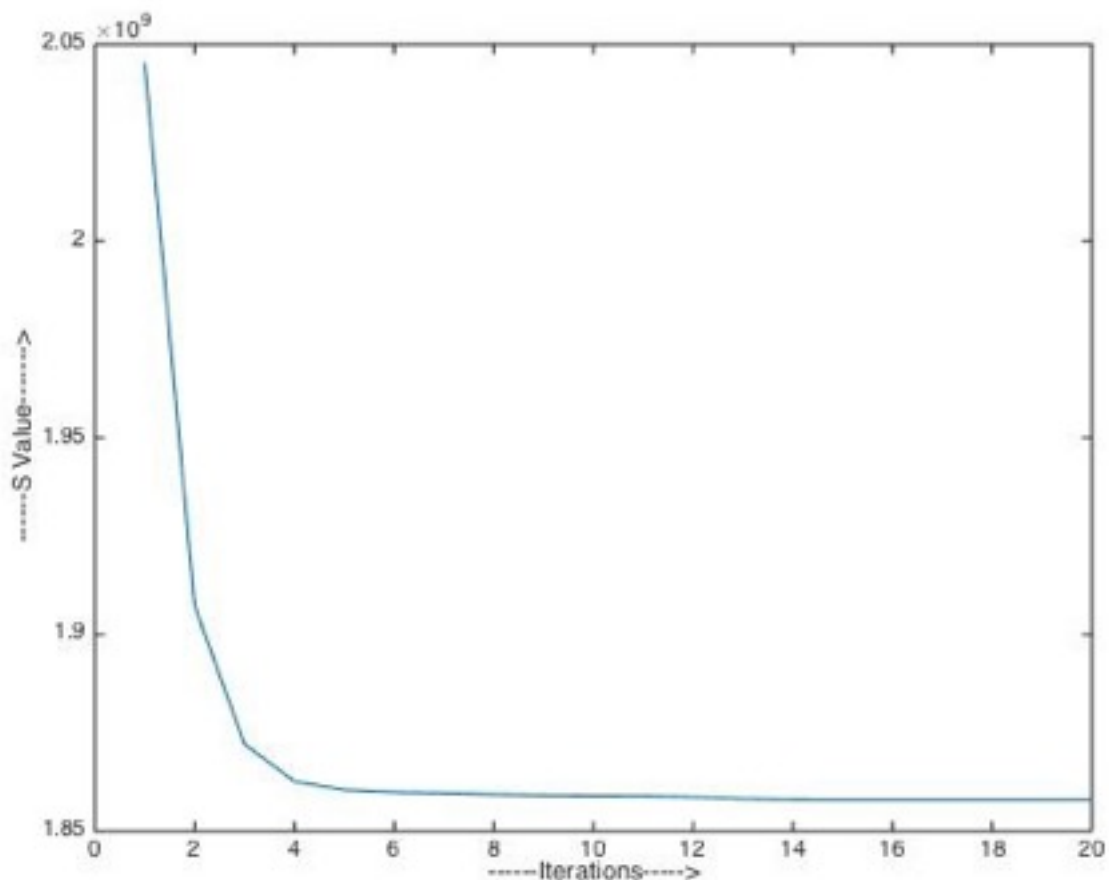
Q1b) The best accuracy obtained is = 79.8%

Number of iterations = 30

The average accuracies range between 75% - 80%

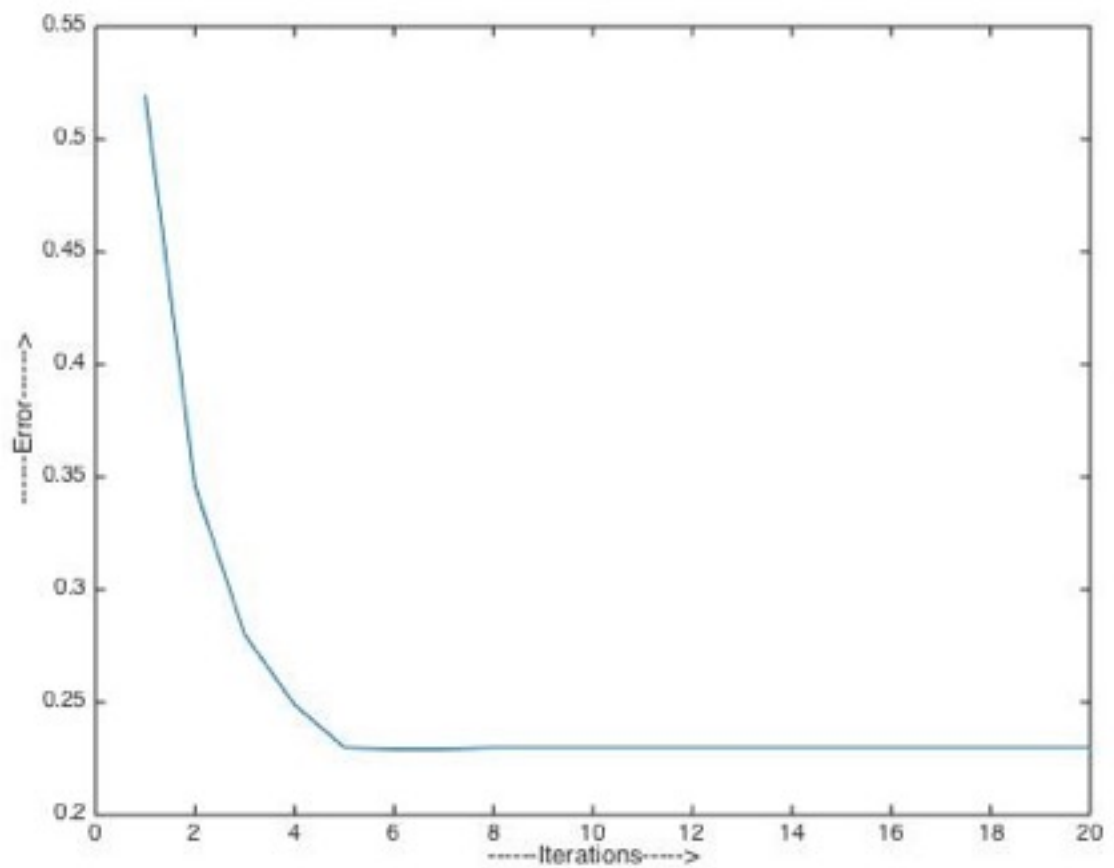
As the means are selected at random there are wide range of accuracies and more than 1 cluster is predicted as to be the same group often.

Q1c)



As the number of iterations increase the quantity S decreases indicating the element is near its mean.

Q1d)



The error gradually decreases with the number of iterations, as the points come closer to actual cluster means.

Q2a)

H=0	H=1
0.804	0.196

	B=0	B=1
H=0	0.9502	0.0498
H=1	0.5837	0.4163

	L=0	L=1
H=0	0.9958	0.0042
H=1	0.7092	0.2908

	X=0	X=1
L=0	0.9779	0.0221
L=1	0.3924	0.6076

	F=0	F=1
B=0, L=0	0.9513	0.0487
B=0, L=1	0.4837	0.5163
B=1, L=0	0.9235	0.0765
B=1, L=1	0.3008	0.6992

log-likelihood = - 2515.3

Q2b) The convergence criterion is that the previous set of parameters and the newly calculated parameters are all equal.

Q2c) **SINGLE MISSING—INITIAL PARAMETERS**

H=0	H=1
0.8018	0.1982

	B=0	B=1
H=0	0.9535	0.0465
H=1	0.5902	0.4098

	L=0	L=1
H=0	0.9952	0.0048
H=1	0.7095	0.2905

	X=0	X=1
L=0	0.9794	0.0206
L=1	0.3949	0.6051

	F=0	F=1
B=0, L=0	0.9521	0.0479
B=0, L=1	0.4930	0.5070
B=1, L=0	0.9093	0.0907
B=1, L=1	0.2717	0.7283

## FINAL PARAMETERS

H=0	H=1
0.8027	0.1973

	B=0	B=1
H=0	0.9529	0.0471
H=1	0.5831	0.4169

	L=0	L=1
H=0	0.9958	0.0042
H=1	0.7128	0.2872

	X=0	X=1
L=0	0.9794	0.0206
L=1	0.3778	0.6222

	F=0	F=1
B=0, L=0	0.9516	0.0484
B=0, L=1	0.4985	0.5015
B=1, L=0	0.9188	0.0812
B=1, L=1	0.2704	0.7296

log-likelihood = -2514.7

## DOUBLE MISSING — —INITIAL PARAMETERS

H=0	H=1
0.8083	0.1917

	B=0	B=1
H=0	0.9455	0.0545
H=1	0.5803	0.4197

	L=0	L=1
H=0	0.9950	0.0050
H=1	0.7034	0.2966

	X=0	X=1
L=0	0.9950	0.0050
L=1	0.7034	0.2966

	F=0	F=1
B=0, L=0	0.9446	0.0554
B=0, L=1	0.3895	0.6105
B=1, L=0	0.9113	0.0887
B=1, L=1	0.2923	0.7077

## FINAL PARAMETERS

H=0	H=1
0.8228	0.1772

	B=0	B=1
H=0	0.9320	0.0680
H=1	0.6065	0.3935

	L=0	L=1
H=0	0.9849	0.0151
H=1	0.7185	0.2815

	X=0	X=1
L=0	0.9789	0.0211
L=1	0.3975	0.6025

	F=0	F=1
B=0, L=0	0.9540	0.0460
B=0, L=1	0.4482	0.5518
B=1, L=0	0.9092	0.0908
B=1, L=1	0.3200	0.6800

log-loglikelihood = -2515.6

The log-likelihood of the previous case and this case are almost similar explaining that the probabilities of the missing values estimated in EM algorithm are correct giving us accurate parameters finally.

The parameters will be more or less the same because they are computed in EM algorithm for several iterations.

### Q3a)

```
1. clf = svm.SVC(C=1.0, class_weight=None, coef0=0.0,  
    decision_function_shape=None, degree=3, gamma='auto', kernel='rbf',  
    max_iter=-1, probability=False, random_state=None, shrinking=True,  
    tol=0.001, verbose=False)
```

Accuracy Test: 2.5e-05

V1 Accuracy: 0.84593333333333

V2 Accuracy: 0.7896

V3 Accuracy: 0.7352666666667

```
2. clf = svm.SVC(C=1.0, class_weight=None, coef0=0.0,  
    decision_function_shape=None, degree=5, gamma='auto', kernel='rbf',  
    max_iter=-1, probability=False, random_state=None, shrinking=True,  
    tol=0.001, verbose=False)
```

V1 Accuracy: 0.84593333333333

V2 Accuracy: 0.7896

V3 Accuracy: 0.7352666666667

```
3. clf = GaussianNB()
```

V1 Accuracy: 0.5388

V2 Accuracy: 0.53973333333333

V3 Accuracy: 0.51923333333333

```
4. clf = svm.SVC(C=1.0)
```

V1 Accuracy: 0.84593333333333

V2 Accuracy: 0.7896

V3 Accuracy: 0.7352666666667

5. clf =  
RandomForestClassifier(criterion="entropy",max\_features=80,min\_samples\_split=5,min\_samples\_leaf=3)  
V1 Accuracy: 0.821533333333  
V2 Accuracy: 0.7391  
V3 Accuracy: 0.6917

6. clf =  
RandomForestRegressor(criterion="mse",max\_features=80,min\_samples\_split=5,min\_samples\_leaf=3)  
V1 Accuracy: 0.499874493113  
V2 Accuracy: 0.326406105373  
V3 Accuracy: 0.222109773045

7. clf = RandomForestRegressor()  
V1 Accuracy: 0.492261935524  
V2 Accuracy: 0.305528663245  
V3 Accuracy: 0.19990530133

8. svm-split3-----c=100  
V1 Accuracy: 0.859666666667  
V2 Accuracy: 0.786866666667  
V3 Accuracy: 0.722766666667  
9. svm-split3-----c=10  
V1 Accuracy: 0.8604  
V2 Accuracy: 0.791133333333  
V3 Accuracy: 0.730666666667

10. svm-split3----c=1,c=10,c=100  
V3 Accuracy: 0.73.2966666667

11. svm-split3----c=5  
V1 Accuracy: 0.852566666667  
V2 Accuracy: 0.792766666667  
V3 Accuracy: 0.7343

12. svm-split3----c=3,c=4,c=5  
V1 Accuracy: 0.846933333333  
V2 Accuracy: 0.791866666667  
V3 Accuracy: 0.7352

13. svm-split5----c=3,c=4,c=5,c=5,c=4  
V1 Accuracy: 0.843566666667  
V2 Accuracy: 0.833533333333  
V3 Accuracy: 0.783366666667  
save2\_split5

14. svm-split5----c=5  
V1 Accuracy: 0.847766666667  
V2 Accuracy: 0.837033333333  
V3 Accuracy: 0.7886  
save5\_split5



```
15. clf =  
DecisionTreeClassifier(criterion='entropy',max_depth=80,min_samples_split=5,min_samples_leaf=  
3)  
V1 Accuracy: 0.7519  
V2 Accuracy: 0.679966666667  
V3 Accuracy: 0.632366666667  
  
16.    clf = svm.LinearSVC(C=1.0)
```

### Report:

1. SVM-Linear: V1 accuracy = 0.5757333  
 V2 accuracy = 0.564466667  
 V3 accuracy = 0.52403333
2. SVM-Gaussian: V1 accuracy = 0.845933333333  
 V2 Accuracy: 0.7896  
 V3 Accuracy: 0.735266666667
3.  
RandomForestClassifier(criterion="entropy",max\_features=80,min\_sampl  
es\_split=5,min\_samples\_leaf=3)  
V1 Accuracy: 0.821533333333  
V2 Accuracy: 0.7391  
V3 Accuracy: 0.6917
4.  
DecisionTreeClassifier(criterion='entropy',max\_depth=80,min\_samples\_sp  
lit=5,min\_samples\_leaf=3)  
V1 Accuracy: 0.7519  
V2 Accuracy: 0.679966666667  
V3 Accuracy: 0.632366666667
5. clf = GaussianNB()  
V1 Accuracy: 0.5388  
V2 Accuracy: 0.539733333333  
V3 Accuracy: 0.519233333333