

## Machine Learning assignment 2 – Report

David VALENSI - 260873585

1. See files in zip
2. GDA model

### a) GDA

GDA performances:

```
The Accuracy is: 0.93
The Precision is: 0.937799043062201
The Recall is: 0.9289099526066351
The F1 Measure is: 0.9333333333333333
```

### b) The learned parameters:

```
w :
[ 14.07043598 -8.63642823 -5.5045595 -3.20634824 -9.60801489
 -4.31571584 16.70297422 -23.54632791 -28.80490041 9.13971621
 -12.88120057 -12.08863337 15.18266729 12.6824781 -5.46015667
 12.79328633 29.05398269 -6.54171831 -0.53833588 -4.7570663 ]

w0 :
26.883504308974533
```

3. a) The k-Nearest Neighbors is much less efficient than GDA:  
The best values of k based on F1 score are generally the even numbers.  
Then, I chose k=8 since it gave the best F1 score about 0.58 (but on other multivariate generated set it could be different k like I got k=4 for another specific set)

### b) k-NN performances on test set are:

```
k: 8
The Accuracy is: 0.54
The Precision is: 0.5594713656387665
The Recall is: 0.6018957345971564
The F1 Measure is: 0.5799086757990868
```

4. See files in zip

## 5. Multiple Multivariate Gaussians

### 1. GDA

a) The achieved performances are:

GDA performances :

```
The Accuracy is: 0.4625
The Precision is: 0.44075829383886256
The Recall is: 0.48947368421052634
The F1 Measure is: 0.4638403990024938
```

b) The learnt parameters:

```
w :
[-0.06230741 -0.01771729 -0.12982761 0.06191243 0.0383158 -0.02825401
-0.08601165 -0.01111571 -0.01205568 0.01182752 0.04258677 0.08571955
0.05765785 0.10556944 -0.08429299 0.00581761 -0.02256758 0.05797166
0.02994211 0.0098346 ]

w0 :
-0.07563267236773956
```

---

2. The k-NN classifiers perform better than GDA on my validation:

```
k: 1
The Accuracy is: 0.515
The Precision is: 0.5196078431372549
The Recall is: 0.5247524752475248
The F1 Measure is: 0.5221674876847291
```

```
k: 2
The Accuracy is: 0.4975
The Precision is: 0.5016393442622951
The Recall is: 0.7574257425742574
The F1 Measure is: 0.6035502958579881
```

```
k: 3
The Accuracy is: 0.53
The Precision is: 0.5339805825242718
The Recall is: 0.5445544554455446
The F1 Measure is: 0.5392156862745098
```

```
k: 4
The Accuracy is: 0.5125
The Precision is: 0.5127272727272727
The Recall is: 0.698019801980198
The F1 Measure is: 0.5911949685534591
```

```
k: 5
The Accuracy is: 0.52
The Precision is: 0.5240384615384616
```

Again, we see that even values for k give the better F1 scores.

3. Best F1 score was achieved with k=2.

Here are the results on test set:

```
k: 2
The Accuracy is: 0.4975
The Precision is: 0.4816053511705686
The Recall is: 0.7578947368421053
The F1 Measure is: 0.5889570552147239
```

6. On a simple Gaussian distribution, the GDA gives very precise performances whereas k-NN classifier gave poorer results around 0.54-0.59.

Nevertheless, when we sample a multiple Gaussian distrusted dataset, it comes that GDA and k-NN classifier both give bad results: around 0.5 for GDA classifier and higher results around 0.55 for k-NN classifier. The k-NN classifier even has a much better F1 measure and recall.

The k-NN classifier is quite constant in term of performances and is less sensitive to the distribution model.