**INF8245E: Machine Learning | Assignment #2**

Louis Plessis (1933334)

24 October 2021

Acknowledgement: I have briefly discussed about this assignment with Joseph Battesti and Marie-Christine Paré.

1. **Linear Classification and Nearest Neighbor Classification**

The dataset can be found in “DS1\_test.txt”, “DS1\_train.txt”, and “DS1\_valid.txt”.

* 1. **GDA model** *(results can vary since the 2000 examples are generated randomly)*
     1. **Best fit accuracy**

Best fit accuracy achieved by the classifier: 0.94875

* + 1. **Learnt coefficients**

Confusion matrix: [[377, 21], [20, 382]]

w = [ 14.25531053 -8.40135958 -5.38105535 -3.26403428 -9.61304149

-4.43431056 16.77321214 -23.84336698 -28.77810374 9.0097474

-13.04880821 -12.49580547 15.55911961 12.55893066 -5.76145043

13.12071568 29.14796768 -6.48674806 -0.3440311 -5.01376722]

w0 = 27.14945783601249

* 1. **K-NN** *(results can vary since the 2000 examples are generated randomly)*

Chart, line chart

Description automatically generated

* + 1. This classifier seems to perform worse than GDA overall (∼0.55 accuracy instead of ∼0.95 for GDA). Some specific values of k seem to perform better than others, due to the low stability of this method. However, the difference is very small as the accuracy stays around 55%.
    2. **Best fit accuracy**

Best fit accuracy: 0.56 when k = 11

* 1. **Mixture of 3 Gaussians**

The dataset can be found in “DS2\_test.txt”, “DS2\_train.txt”, and “DS2\_valid.txt”.

* 1. **DS2** *(results can vary since the 2000 examples are generated randomly)*
     1. **GDA model**
        1. **Best fit accuracy**

Best fit accuracy achieved by the classifier: 0.6225

* + - 1. **Learnt coefficients**

Confusion matrix: [[232, 149], [153, 266]]

w = [-0.06865608 0.04504302 -0.07562544 0.01710506 0.27634092 -0.06246668

-0.26920294 0.01491341 0.21412222 0.10008398 0.16090591 0.10046235

0.13253823 0.0796615 -0.19677276 0.0210705 -0.08417598 0.02937124

-0.1133954 -0.00365615]

w0 = -0.462974592329536

* + 1. **K-NN**

Chart, line chart, histogram

Description automatically generated

This classifier seems to perform worse than GDA overall (∼0.45 accuracy instead of ∼0.95 for GDA). Some specific values of k seem to perform better than others, due to the low stability of this method. However, the difference is very small as the accuracy stays around 45%.

* + 1. **Best fit accuracy with K-NN**

Best fit accuracy: 0.43 when k = 2

* 1. **Similarities and differences between the performance of both classifiers**

With both DS1 and DS2, GDA seems to perform better than k-NN. Also, k-NN behaves in the same way with both datasets (i.e., some values of k perform better than others, with no visible pattern). However, best fit accuracy for k-NN is found earlier with DS2 (low k value) than with DS1 (𝑘∼30). Besides, DS1 seems to perform better than DS2 overall (∼95% and ∼55% accuracies for DS1, ∼62% and ∼45% accuracies for DS2).

1. **MNIST Handwritten Digits Classification**
   1. **GNB model**

Data preprocessing steps can be found in the Jupyter Notebook.

* + 1. **Equations for mean and diagonal covariance matrices**

X

* + 1. **Estimating GNB model parameters**

X

* 1. **K-NN**

X

* + 1. Bla

X

* + 1. **Best fit accuracy**

X

* 1. **GNB performance vs k-NN**

X