**COMP-551 : Applied Machine Learning Given on: Jan 30, 11pm Programming Assignment #2 Due on : Jan 12, 12pm**

Student’s name: Frederic Ladouceur Student’s ID: 260690491

1. You can find the synthetic datasets *DS1*, *DS1*\_*train* and *DS1*\_*test* under the path: out\A2.1\<file\_name>.csv
2. The best fit produced the following results:

* Accuracy: 0.960833333333

Precision: 0.961279461279

Recall: 0.959663865546

F\_measure: 0.96047098402

* Coefficients estimated for class 0 are:

w0= 26.09831444

w= [ 13.83781553 -8.13577882 -5.2890129 -3.11571921 -9.16358703

-4.34955068 16.03045348 -22.79275583 -27.57560007 8.69481137

-12.67621349 -11.75712713 14.65557009 12.17681873 -5.55571079

12.53713866 28.09403341 -6.37317519 -0.43307619 -4.62221076]

* Coefficients estimated for class 1 are:

w0= -26.09831444

w= [-13.83781553 8.13577882 5.2890129 3.11571921 9.16358703

4.34955068 -16.03045348 22.79275583 27.57560007 -8.69481137

12.67621349 11.75712713 -14.65557009 -12.17681873 5.55571079

-12.53713866 -28.09403341 6.37317519 0.43307619 4.62221076]

1. This k-NN classifier model performs poorly compared to the LDA model. The performance measures give us a clever way to compare both algorithms, and the LDA model produces significantly better results. Below are the performance measures for the k-NN model that produce the best fits:

* Accuracy: 0.589166666667 Number of neighbours (k): 340
* Precision: 0.570502431118 Number of neighbours (k): 340
* Recall: 0.606896551724 Number of neighbours (k): 340
* F\_measure: 0.58813700919 Number of neighbours (k): 340

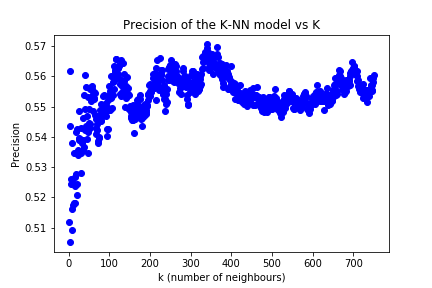
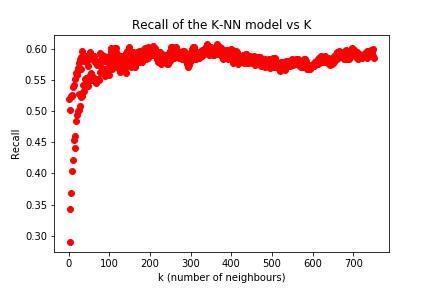
The following graphs show how the different performance measures vary with the number of neighbours (k) on DS1.

Figure : F\_measure vs K on DS1

Figure : Recall vs K on DS1

Figure : Precision vs K on DS1

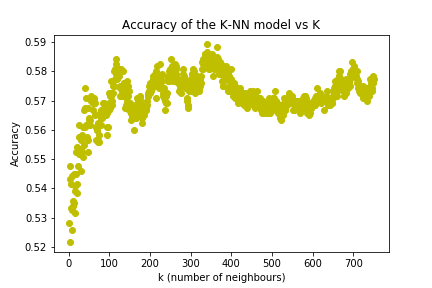
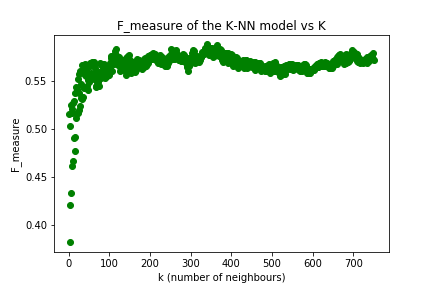


Figure : Accuracy vs K on DS1

1. You can find the synthetic datasets *DS2*, *DS2*\_*train* and *DS2*\_*test* under the path: out\A2.4\<file\_name>.csv

*LDA Classifier*

The best fit of the LDA classifier produced the following results:

* Accuracy: 0.5275

Precision: 0.54351687389

Recall: 0.496753246753

F\_measure: 0.519083969466

* Coefficients estimated for class 0 are:

w0= 0.07782143

w [ 0.03095733 0.00568334 0.04279177 0.05273383 -0.04133535 0.09384691

0.0326761 -0.02106459 -0.03836051 -0.09044405 -0.01191598 -0.01282974

-0.03690949 -0.01060537 0.01106433 -0.01245718 0.06254451 -0.01113789

-0.01637733 -0.07398869]

Coefficients estimated for class 1 are:

w0= -0.07782143

w= [-0.03095733 -0.00568334 -0.04279177 -0.05273383 0.04133535 -0.09384691

-0.0326761 0.02106459 0.03836051 0.09044405 0.01191598 0.01282974

0.03690949 0.01060537 -0.01106433 0.01245718 -0.06254451 0.01113789

0.01637733 0.07398869]

LDA observations: The performance measures of the LDA classifier on this dataset are significantly lower than the results we had in 2.2. It indicates us that the LDA classifier does not perform as good on DS2 as it did on DS1.

*k-NN Classifier*

* Accuracy: 0.561666666667 Number of neighbours (k): 12
* Precision: 0.591939546599 Number of neighbours (k): 9
* Recall: 0.50974025974 Number of neighbours (k): 4
* F\_measure: 0.537785588752 Number of neighbours (k): 12

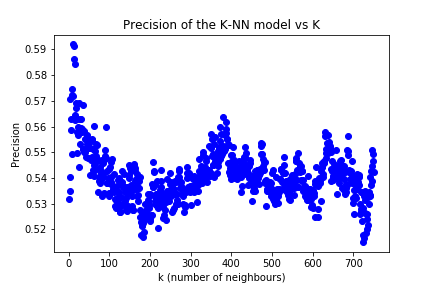
k-NN Observation: This k-NN classifier on DS2 performs as poorly as it did with DS1. One thing we can notice is that best performance is achieved with fewer neighbours on DS2 than on DS1. Also, the performance of the classifier seems to decrease more rapidly (with k) on DS2 than on DS1. The performance measures above are low which indicates poor performance. The following graphs show how the different performance measures vary with the number of neighbours (k) on DS2.

Figure : Precision vs K on DS2

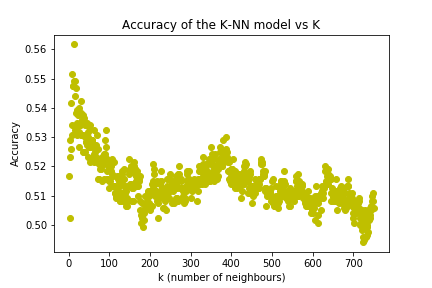


Figure : Accuracy vs K on DS2

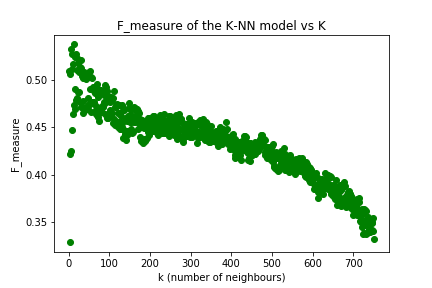


Figure : F\_measure vs K on DS2

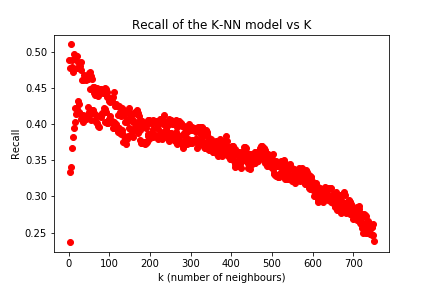


Figure : Recall vs K on DS2

General observations: Both, the LDA and the k-NN classifier perform poorly on DS2 as shown by the performance measures. However the k-NN performs slightly better than the LDA classifier.

The LDA classifier performs better on DS1 than on DS2 according to the performance measures. The best fit parameters found for DS1 are also larger than the ones found for DS2. We understand that the data in DS1 are well suited for a linear classifier.

The k-NN performs slightly better on DS1 than on DS2. The best choice of k (number of neighbours) hyperparameter is smaller in DS2 than on DS1. Also, the performance measures decrease faster (w.r.t. k) when using DS2 than DS1. None of these datasets are “locally smooth” enough to be classified using a k-NN algorithm.