|  |
| --- |
| **Exercise 1: Learning in discrete graphical models**    with N the size of the sample. |
|  |
| **Exercise 2.1.(a): LDA formulas**  with N the size of the sample and  Conditional probability (it has the same form than logistic regression!):  With: |
|  |
| **Exercise 2.5.(a): QDA formulas**  Conditional probability:  With: |

**Dataset A**

|  |  |
| --- | --- |
| **/var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/A1AB5FA0.tmp**  LDA | **/var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/438C97DA.tmp**  Logistic |
| **/var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/8600DDC4.tmp**  Linear regression | /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/2AD84550.tmp  QDA |
| **Table of misclassifications for test set**   |  |  |  |  | | --- | --- | --- | --- | | LDA | Logistic | Linear | QDA | | 2.87 % | 2.07 % | 2.07% | 1.93% | | LDA appears being highly unproductive as the dataset is ill-conditionned. All of the other ones provide good results with correct generalization (test set’s performances are close to train set’s), as the points are linearly separable. |

**Dataset B**

|  |  |
| --- | --- |
| /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/9B9906AE.tmp  LDA | /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/EC2BFCF8.tmp  Logistic |
| **/var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/8600DDC4.tmp**  Linear | /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/5B47E7DE.tmp  QDA |
| **Table of misclassifications for test set**   |  |  |  |  | | --- | --- | --- | --- | | LDA | Logistic | Linear | QDA | | 5.00 % | 3.80 % | 4.15% | 3.45% | | LDA provides bad results, because the assumption is not respected. appears being highly unproductive as the dataset is ill-conditionned. The logistic and linear regression provide modest results, because the data positive and negative overlap. QDA provides best results, as it can exclude points finer. |

**Dataset C**

|  |  |
| --- | --- |
| /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/CE9556EC.tmp  LDA | /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/564C85C6.tmp  Logistic |
| **/var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/8600DDC4.tmp**  Linear | /var/folders/ty/1lsvbm5d23jfrmrt1cqkzy080000gn/T/com.microsoft.Word/Content.MSO/E53E3C72.tmp  QDA |
| **Table of misclassifications for test set**   |  |  |  |  | | --- | --- | --- | --- | | LDA | Logistic | Linear | QDA | | 2.7 % | 2.87 % | 4.23% | 3.87% | | Prior models are wrong, because the positive samples are acting like a mixture of 2 Gaussians. Hence, LDA and QDA provide wrong estimate of variance (blue ellipses). The least square of linear regression provide a too important weight to the bottom left blob. Only the logistic is robust to this perturbation. |

**Proof**

We consider only the case of the QDA. To get the proof of LDA, we consider the case where

The log-likelihood is defined by:

By definition of the probability density, we can separate the log likelihood into two different likelihood and maximize both of them:

The two likelihood are concave with respect to their parameters. Consequently, their maximum values are achieved when the gradient is zero.

Using the gradient of and as explained in lessons, we have:

Finally, to get the boundary, we use the Bayes rule:

Replacing the definition of the density function, we have the following equivalence:

With: