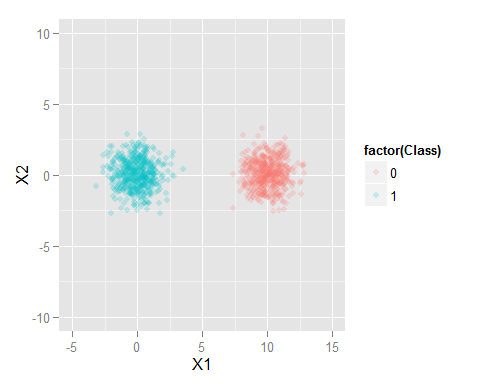
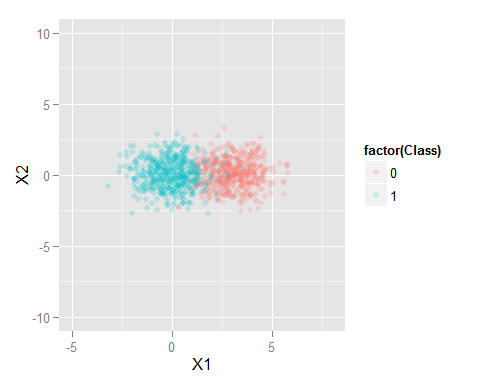
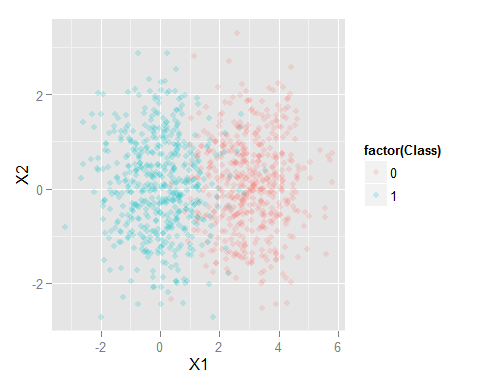
**CSCE 623 Spring 2020 - Machine Learning: In Class Work, Day 8**

From Chapter 4: Classification

1. Essay: Why are Logistic Regression’s training set parameters *unstable* when the classes’ true populations have a wide separation in their high-density regions? Hint – think about what *unstable* means for the values of parameters that can fit the data when the training set is just a small sample of the true population. These figures may help:



1. Consider the following Confusion Matrices which were generated from a single classifier on a single dataset (shown right), using different probability cutoff thresholds.
   1. Compute: Accuracy, False Positive Rate (FPR), True Positive Rate (TPR; Recall), Precision and F-measure (F1 score), for each matrix (*hint – make a function*). Note the probability threshold listed in the top left corner will be used for part b and c.   
       p=predicted, a=actual, P=positive, N=Negative

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0.0001 | aP | aN |  | ACC: |
| pP | 27 | 0 |  | FPR: |
| pN | 218 | 255 |  | TPR(recall): |
|  |  |  |  | Precision: F-measure: |
|  |  |  |  |  |
|  |  |  |  |  |
| 0.25 | aP | aN |  | ACC: |
| pP | 219 | 10 |  | FPR: |
| pN | 26 | 245 |  | TPR(recall): |
|  |  |  |  | Precision: F-measure: |
|  |  |  |  |  |
| 0.5 | aP | aN |  | ACC: |
| pP | 233 | 17 |  | FPR: |
| pN | 12 | 238 |  | TPR(recall): |
|  |  |  |  | Precision: F-measure: |
|  |  |  |  |  |
|  |  |  |  |  |
| 0.75 | aP | aN |  | ACC: |
| pP | 243 | 39 |  | FPR: |
| pN | 2 | 216 |  | TPR(recall): |
|  |  |  |  | Precision: F-measure: |
|  |  |  |  |  |
|  |  |  |  |  |
| 0.9999 | aP | aN |  | ACC: |
| pP | 245 | 236 |  | FPR: |
| pN | 0 | 19 |  | TPR(recall): |
|  |  |  |  | Precision: F-measure: |

1. Plot the 5 TPR/FPR points on the ROC graph and draw the estimated ROC curve
2. Estimate the area under the curve (AUC).

1.0



False Positive Rate

0.0

0.5

0.5

0.0

True Positive Rate

1.0