Figure 1: Effect of the number of clusters on the sum of square error for the mushroom (left) and the pulsar star (right) dataset

Figure 2: Effect of the number of clusters on the log likelihod for the mushroom and the pulsar star dataset

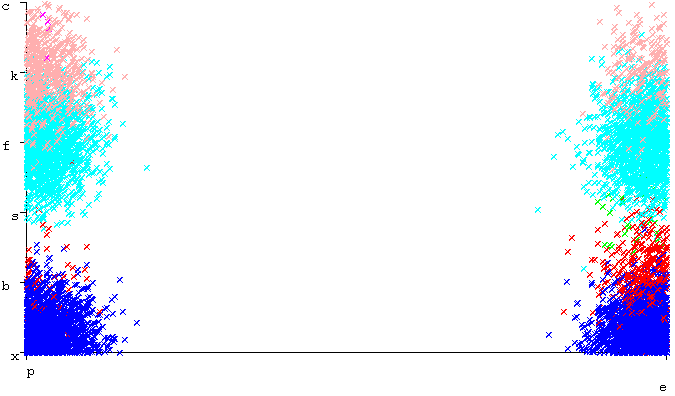


Figure 3: Raw mushroom data looking at habitat on the left and cap shape on the right with respect to their classification

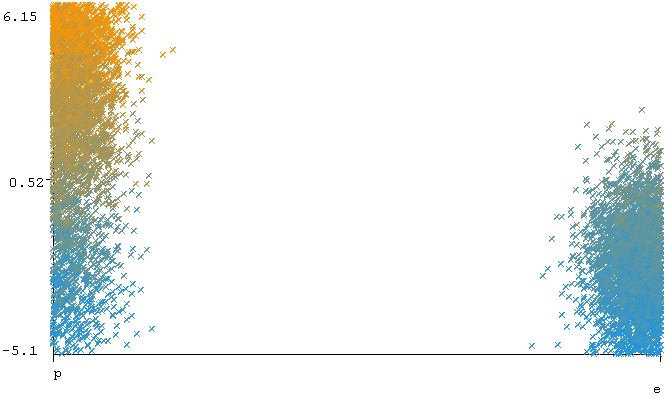
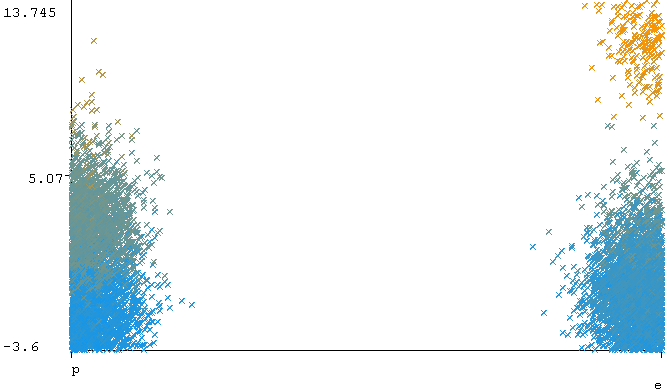


Figure 4: Mushroom data after PCA with 2 new attributes

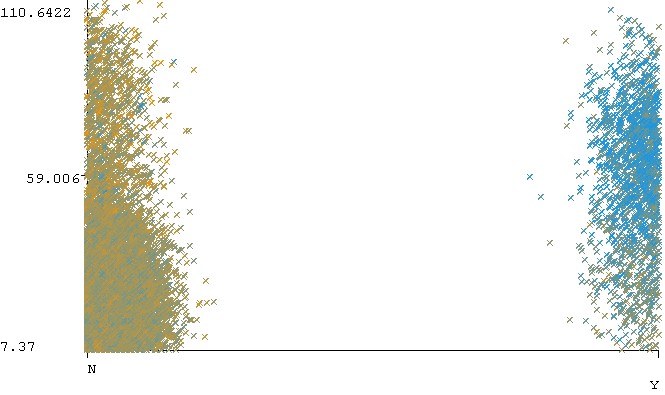
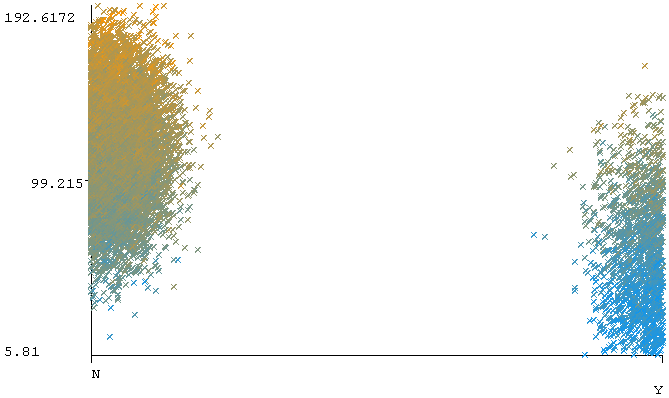
 

Figure 5: Raw pulsar star data looking at standard deviation of the DM-SNR curve on the left and mean of integrated profile on the right with respect to their classification

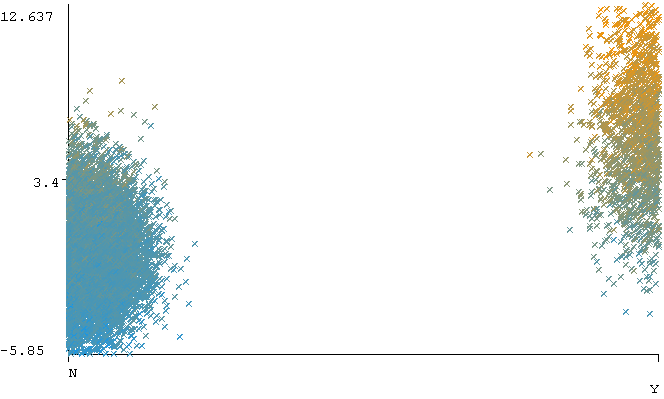
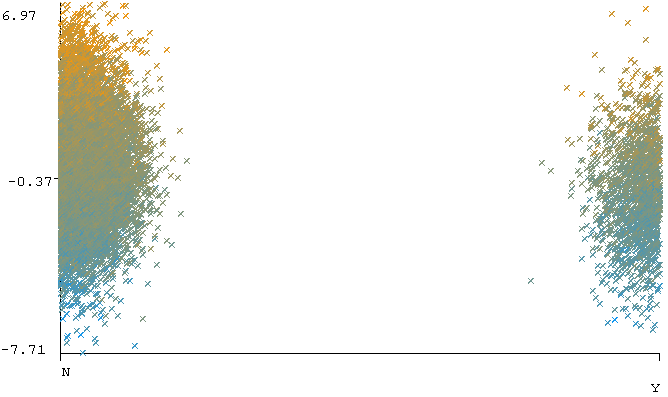
 

Figure 6: Pulsar star data after PCA with 2 new attributes

|  |  |
| --- | --- |
| Attribute | EigenValue |
| Mushroom New Attribute 1 | 7.78 |
| Mushroom New Attribute 2 | 9.84 |
| Pulsar Star New Attribute 1 | 4.13 |
| Pulsar Star New Attribute 2 | 2.14 |

Table 1.

Figure 7: Effect of the number of clusters on the sum of square error for the mushroom and the pulsar star dataset after using PCA

Figure 8: Effect of the number of clusters on the log likelihood for the mushroom and the pulsar star dataset after using PCA

|  |  |
| --- | --- |
| NN on original mushroom | NN on PCA mushroom |
| NN on mushroom with k-means | NN on mushroom with EM |

|  |  |
| --- | --- |
| NN on original mushroom | NN on PCA mushroom |
| NN on mushroom with k-means | NN on mushroom with EM |