Discriminative vs. Generative classifiers

A comparison of perceptron and naive Bayes

Introduction

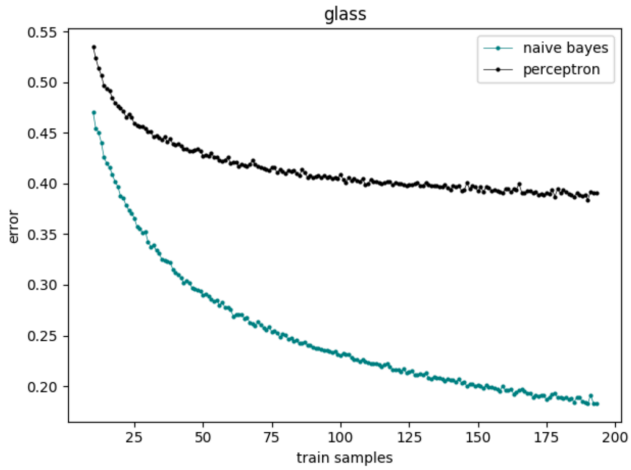
Based on the paper by Andrew Y. Ng and Michael I. Jordan *Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes* (2002) I tried to show that a discriminative classifier performs better with a bigger dataset as it has a lower asymptotic error, but unlike their article, I work with perceptron instead of logistic regression. On the other hand the generative classifier has a faster learning rate, performing better when the dataset contains a few elements as it reaches its higher asymptotic error faster.

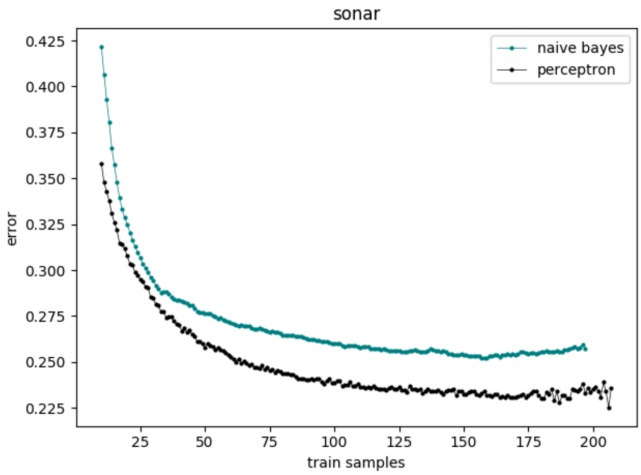
Preliminaries

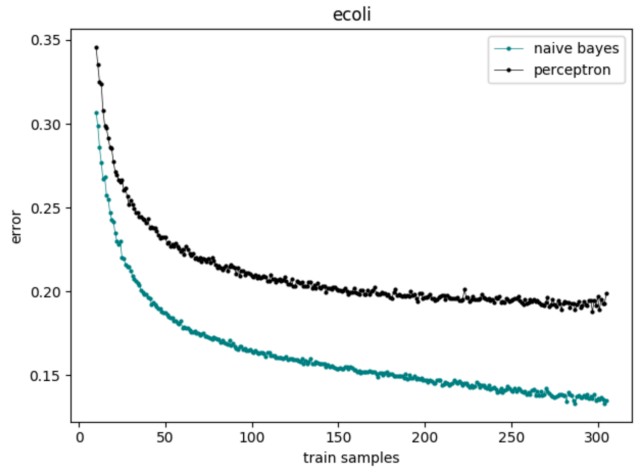
To run the experiments I used libraries provided by scikit-learn.
I used two kinds of naive Bayes classifier: gaussian and multinomial. The first one fits continuous values while the second one performs better on categorical values.
The constructor of GaussianNB takes as parameter v*ar_smoothing*, which is the portion of the largest variance of all features that is added to variances for calculation stability. In some cases, the increasing of this number helps to achieve a faster learning rate.
The constructor of MultinomialNB takes as parameter *alpha,* the additive (Laplace) smoothing parameter, increasing this number helps especially in small datasets where there are a few/no occurrences of some attributes.
Regarding the perceptron classifier there are two knobs:
max_iter: the maximum number of epochs. Increasing this number may help to reach a lower error combined with the *eta0* parameter.
eta0: constant by which the updates are multiplied. Decreasing this number might result into finding more accurate bounds, especially when the dataset is not linearly separable.

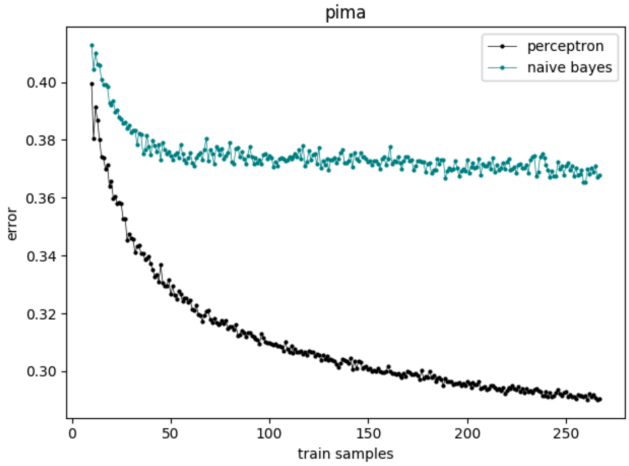
Experiments

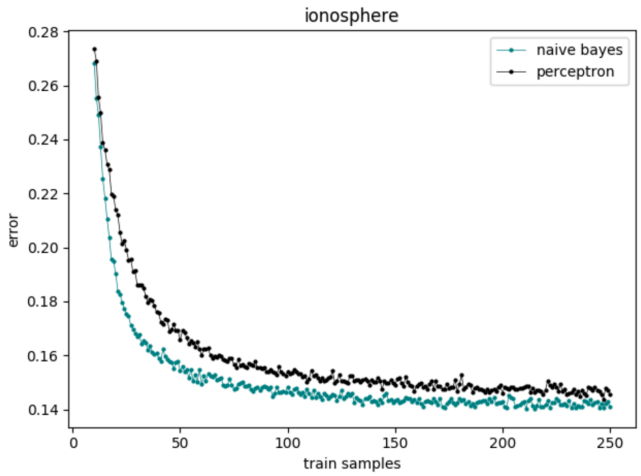
The following results were obtained from datasets with only continuous values.



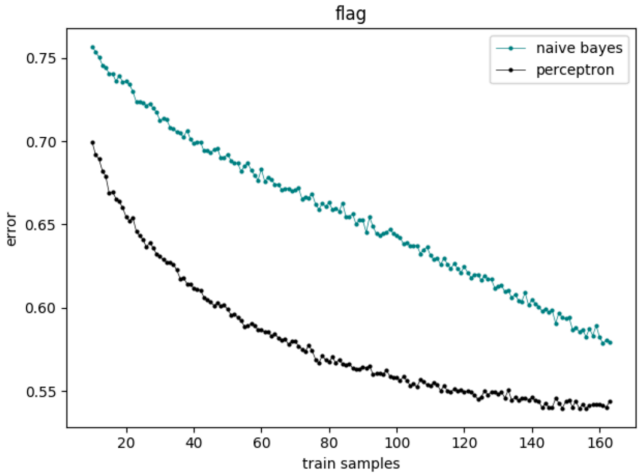


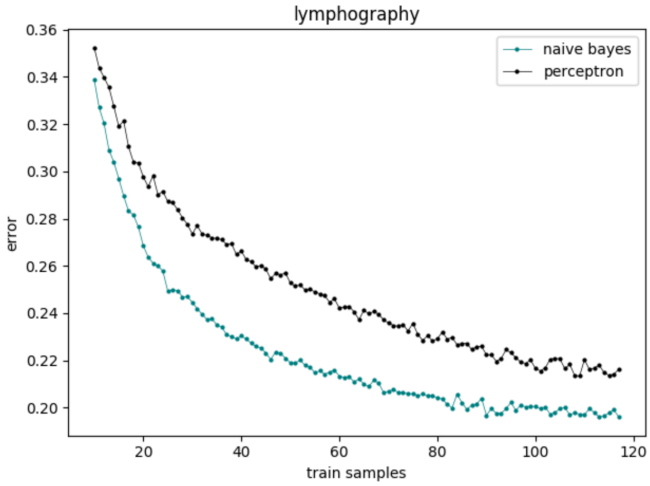


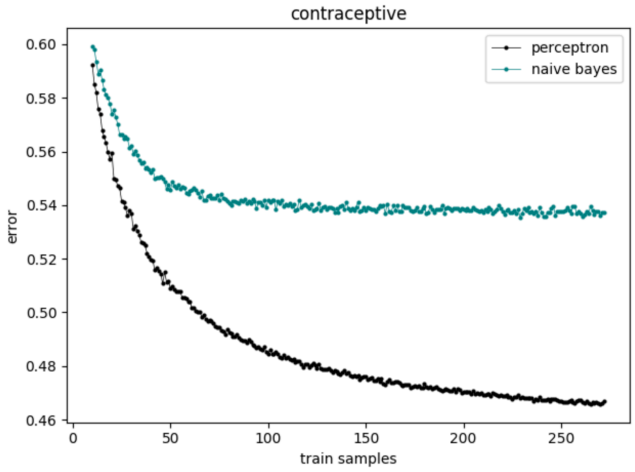


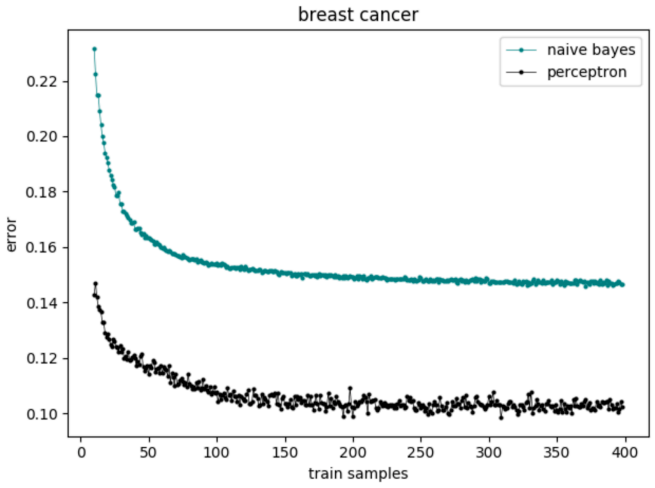


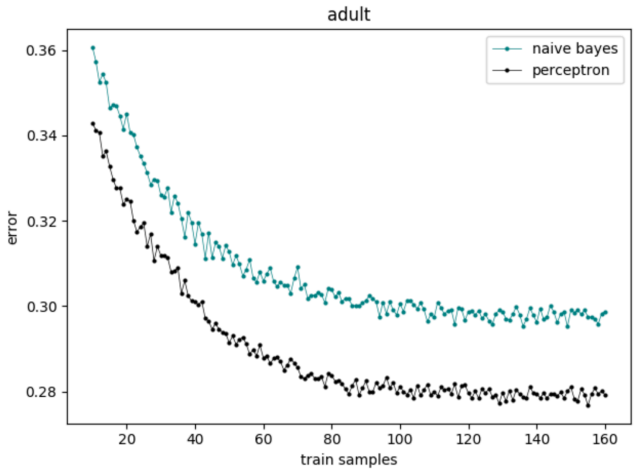
The following results where obtained from datasets with only categorical values.











Conclusion

Using the implementations made by scikit-learn I haven't seen a considerable difference in the performance of the two algorithms, as some support the thesis and some don't, even after repeated improvements on the parameters of the classifiers. Therefore my results are different from those of the article, but it is still important to consider that the algorithms used in that paper are significantly different from the one used in these experiments.