Naive Bayes Classifiers

Similar to Linear Models but with the advantage of faster training and the disadvantage of relatively worse generalization performance than Linear Classifiers like `LogisticRegression` and `LinearSVC`. This efficiency is due to the way NBCs learn parameters by looking at each feature individually and collecting simple per class stats from each feature. There are three kinds of NBCs implemented in the scikit-learn package:

- `GaussianNB`
- `BernoulliNB`
- `MultinomialNB`

BernoulliNB Binary data, mostly used in text data classification. Counts how often every feature of each class is not zero.

MultinomialNB Count data (each feature represents and integer count of something e.g. how often a word appears in a sentence), mostly used in text data classification. Takes into account average value of each feature of each class.

GaussianNB Continuous data. Stores the average value as well as the standard deviation of each feature for each class

Strengths, Weaknesses and Parameters

Strengths

- Very fast to train

- Training procedure is easy to understand

- Works well with high-dimensional sparse data

- Used as baseline models on very large datasets where training even a linear model would take too long

- `MultinomialNB` and `BernoulliNB` - Used for sparse count data such as text. Multinomial usually performs better particularly on datasets with a relatively large number of nonzero features e.g. large documents

- `GaussianNB` - Mostly used on very high-dimensional data.

Weaknesses

Parameters

- `Alpha` - The only parameter `MultinomialNB` and `BernoulliNB` have. Controls model complexity. The way alpha works is that the algorithm adds to the data "`alpha`" many virtual data points that have positive values for all the features, resulting in a smoothing of the stats. A large `alpha` value means more smoothing resulting in less complex models.