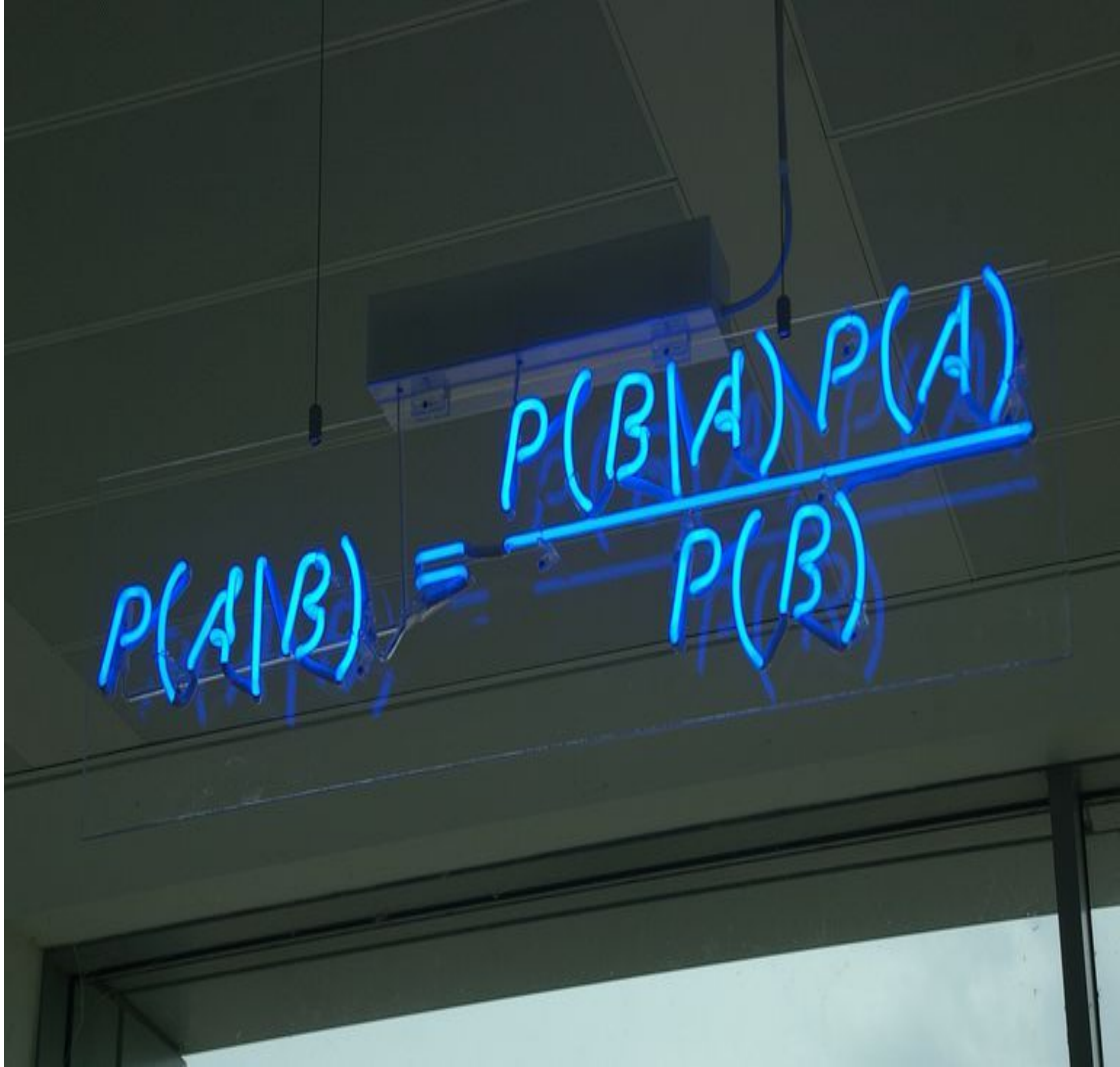
NAIVE BAYES CLASSIFIER



Algorithms

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Estimating the parameters

$$\begin{aligned} c &= \arg\max_{c} p(c|\mathbf{x}) \\ &= \arg\max_{c} \frac{p(\mathbf{x}|c)p(c)}{p(\mathbf{x})} \\ &= \arg\max_{c} p(\mathbf{x}|c)p(c) \end{aligned}$$

With:

$$p(\mathbf{x}|c) = p(x_1, x_2, \ldots, x_d|c) = \prod_{i=1}^{d} p(x_i|c)$$

Note: p(c) can be calculated using **MLE** or **MPA** estimation, the first is commonly used.

Variations

- **Multinomial Naive Bayes:**

  Mostly used for document classification problem, i.e whether a document belongs to the category of sports, politics, technology etc. The features/predictors used by the classifier are the frequency of the words present in the document.

- **Bernoulli Naive Bayes:**

  This is similar to the multinomial naive bayes but the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

- **Gaussian Naive Bayes:**

  When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a gaussian distribution.

Loss function

Maximum Log-likelihood or Negative Log-likelihood

$$\theta = \max_{\theta} \sum_{n=1}^{N} \log(p(\mathbf{x}_n|\theta))$$

Overfitting

- When: Not likely, if properly implemented.

- To avoid overfitting:
  - Use k-fold cross validation
  - Have a validation dataset

References:

1. https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c
2. https://machinelearningcoban.com/2017/08/08/nbc/
3. https://machinelearningcoban.com/2017/07/17/mlemap/#-maximum-likelihood-estimation
4. https://www.slideshare.net/ananth/an-overview-of-nave-bayes-classifier
5. https://www.saedsayad.com/naive_bayesian.htm
6. https://stackoverflow.com/questions/25583591/loss-risk-function-for-sci-kit-learns-naive-bayes-classifier
7. https://stats.stackexchange.com/questions/296014/why-is-the-naive-bayes-classifier-optimal-for-0-1-loss
8. https://stats.stackexchange.com/questions/141087/i-am-wondering-why-we-use-negative-log-likelihood-sometimes
9. https://www.quora.com/Why-are-Naive-Bayes-classifiers-considered-relatively-immune-to-overfitting
10. https://medium.com/@vibhuti.siddhpura/machine-learning-algorithms-introduction-fb86623c5218