Naive Bayes Classifier

The purpose of this is to test how well the naive Bayes model fits the data in contrast to the one used in the research paper.

But first we need to define our model. A Bernoulli naive Bayes is good for the purpose: it uses binary data to sort samples into categories; in this case two. The model has the following structure

$$p(C_i \mid \mathbf{x}) \propto p(\mathbf{x} \mid C_i)p(C_i) = p(C_i)\Pi_{j=1}^{N}p(x_j \mid C_i)^{b_j}(1 - p(x_j \mid C_i))^{(1-b_j)}$$

where b_j denotes the boolean value of having the TCR of the index j.

Moreover, these probabilities are based on the expected values of the proportions in the sample. This means that $p(x_j \mid C_i)$ is actually $p(x_j \mid C_i, \mathbf{x}_0)$ in this case. Due to the independence assumption, we can first calculate the expected values of these, and then calculate the product.

These proportional values are then transformed into probabilities

$$p(C_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_i)p(C_i)}{\Sigma_{n=1}^{K}p(\mathbf{x} \mid C_n)p(C_n)}$$

Source:

```
# here python code if necessary
```