

## 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

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

where  $b_j$  denotes the boolean value of having the TCR of the index  $j$ . These proportional values are then transformed into probabilities

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

**Source:**

```
import random
from math import log
from math import exp

# plot that shit
import matplotlib.pyplot as plt
plt.hist(s, bins='auto')
plt.title("Samples from Exp(2) ")
plt.show()
```