Bayes Theorem:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

Y: class label
X: features
P(Y): prior probability
P(X|Y): Likelihood
P(X): Overall probability of the feature
P(Y|X): posterior probability

Simple Example of Bayes theorem:

Assume:
- 40 percent of all emails are spam.
- 20 percent of spam emails contain the word lottery.
- 2 percent of non spam emails contain the word lottery

Suppose you receive an email that contains the word "lottery".
You want to calculate: P(spam | lottery)

Let,
Lottery = L
Spam = S
Not Spam = NS

Given,
P(S) = 0.40
P(NS) = 0.60
P(L|S) = 0.2
P(L|NS) = 0.02

P(S|L) = ?

$$P(S|L) = \frac{P(L|S) \cdot P(S)}{P(L)}$$

$$0.2 \times 0.40$$

P(L) = ?

$$P(L) = P(L|S) \times P(S) + P(L|NS) \times P(NS)$$

$$0.2 \times 0.40 +$$

$$= \frac{0.2 \times 0.40}{0.092}$$

$$= 0.87$$

87%

$$+ P(\quad)$$

$$= 0.2 \times 0.40 + 0.02 \times 0.60$$

$$P(2) = 0.092$$

The Naive Assumption:

For features $x_1, x_2, \ldots x_n$

$$P(x_1, x_2, \ldots x_n \mid Y) = \prod_{i=1}^{n} P(x_i \mid Y)$$

Example:

three words in an email

x_1 = offer, x_2 = free, x_3 = win

$$P(x_1, x_2, x_3 \mid \text{spam}) = P(\text{offer} \mid \text{spam}) \times P(\text{free} \mid \text{spam}) \times P(\text{win} \mid \text{spam})$$

Training and Prediction Mechanics:

1. Goal of Naive Bayes

$$\hat{Y} = \arg\max_{c_k} P(Y = c_k \mid X)$$

2. Training Step:

a) priors:

$$P(Y = c_k)$$

b) Likelihood:

$$P(X_i \mid Y) =$$

n of spam email containing free

$P(X_i|Y) =$

$\Downarrow$

$P(\text{free}=1|\text{spam}) = \frac{\text{n of spam email containing free}}{\text{number of spam email}}$

$= \frac{0}{X}$

Laplace smoothing?

$P(X_i|Y) = \frac{\text{count}+1}{\text{total}+k} = \frac{1}{}$

3. Prediction Step:

if features are: $x_1, x_2, \ldots x_n$

Compute for each class:

$$\text{score}(c_k) = P(Y=c_k)\prod_{i=1}^{n} P(X_i|Y=c_k)$$

Simple Example of Prediction

Our email has offer and free word in it.

Assume,

$P(\text{spam}) = 0.4$

$P(\text{not spam}) = 0.6$

$P(\text{offer}|\text{spam}) = 0.2$

$P(\text{free}|\text{spam}) = 0.10$

$P(\text{offer}|\overset{\text{not}}{\sim}\text{spam}) = 0.02$

$P(\text{free}|\text{not spam}) = 0.01$

Compute spam score:

$\text{score}(\text{spam}) = 0.40 \times 0.2 \times 0.10$

$= 0.008$

Compute not spam score:

$\text{score}(\text{not spam}) = 0.60 \times 0.02 \times 0.01$

$= 0.00012$

$0.008 > 0.00012$

$0.008 > 0.00012$

The email is spam