



MACHINE LEARNING NAÏVE BAYES CLASSIFIER

Brijesh Bhatt
17 July 2020



Naïve Bayes Classifier

- **classification** is the problem of identifying to which of a set of **categories** (sub-populations) a new **observation** belongs, on the basis of a **training set** of data containing observations (or instances) whose category membership is known.
- **probabilistic classifier** is a **classifier** that is able to predict, given an observation of an input, a **probability distribution** over a **set** of classes, rather than only outputting the most likely class that the observation should belong to.
- Naïve Bayes classifier is a probabilistic classifier developed based applying Bayes theorem with independence assumption between features.
- Naïve Bayes classifier is a generative model.

Are you infected by Covid -19?

- Covid infects around 30% of the population.
- 70% of people who are tested are found positive.
- You are tested positive with an overall accuracy of testing = 80%

Naïve Bayes classifier formulation

The diagram shows the equation for the Naïve Bayes classifier. The equation is $P(c | x) = \frac{P(x | c)P(c)}{P(x)}$. Four blue arrows point from text labels to parts of the equation: 'Likelihood' points to $P(x | c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c | x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Labels with arrows:

- Likelihood (points to $P(x | c)$)
- Class Prior Probability (points to $P(c)$)
- Posterior Probability (points to $P(c | x)$)
- Predictor Prior Probability (points to $P(x)$)

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

Are you infected by Covid-19?

X1 Contact	X2 Symptom	X3 Test	C infected
Yes	Yes	Yes	Yes
Yes	Yes	No	Yes
Yes	Yes	No	No
No	Yes	Yes	Yes
No	No	Yes	No

X1 Contact	X2 Symptom	X3 Test	C infected
Yes	No	Yes	Yes
Yes	No	No	Yes
No	Yes	Yes	Yes
Yes	No	Yes	Yes
Yes	No	No	Yes

Are you infected?

- You are not in the contact of a covid positive, you have symptoms, you are tested negative,
- You are not in contact of a covid positive, you have no symptoms and you are tested negative.

Conditional probability table

$P(x_1 = T)$

$P(X_2 = T)$

$P(X_3 = T)$

$P(C = T)$

$X_1 \backslash C$	T	F
T		
F		

$X_2 \backslash C$	T	F
T		
F		

$X_3 \backslash C$	T	F
T		
F		

$P (C = T / X1 = F, X2 = T, X3 = F)$

$$P (C = T / X_1 = F, X_2 = F, X_3 = F)$$