Naive Bayes Classifier: solved example

* Probabilistic classifier.
* generative model.
* calculates the joint distribution p (X, c), where X is the feature set, c is the class label.
* the Naive Bayes conditional independence assumption, assume that the attribute values are independent of each other given the class.
* probability values can be calculated quickly. Usually, very fast algorithm.
* Learning : Learning parameters : class prior, likelihood.
  The algorithm learns prior p(c) and likelihood p(xi/c) from the data.
* Prediction / inference : using the likelihood and prior (calculated in learning), the algorithm calculates posterior probability by applying Bayes rule.

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
Yes	No	Yes	Yes
Yes	No	No	Yes
No	Yes	Yes	Yes
Yes	No	Yes	Yes
Yes	No	No	Yes

$$P(c=Yes) = \frac{\#\text{ Examples with } c=Yes}{\text{Total }\#\text{ Examples}}$$

$$P(X_i = Yes / c = Yes) = \frac{\#\text{ Examples with } X_i=Yes \;\&\; c=Yes}{\#\text{ Examples with } c = Yes}$$

* Learning : Conditional Probability tables

(A) Class Prior

$P(c = Yes) = \frac{8}{10}$; $P(c = No) = \frac{2}{10}$

c	Yes	No
	8/10	2/10

← class prior

(B) Likelihood : $P(X_i / c)$

$P(x_1=\text{Yes}/c=\text{Yes}) = \frac{6}{8}$; $P(x_1=\text{No}/c=\text{Yes}) = 2/8$

$P(x_1=\text{Yes}/c=\text{No}) = 1/2$; $P(x_1=\text{No}/c=\text{No}) = 1/2$

$P(x_2=\text{Yes}/c=\text{Yes}) = 4/8$; $P(x_2=\text{No}/c=\text{Yes}) = 4/8$

$P(x_2=\text{Yes}/c=\text{No}) = 1/2$; $P(x_2=\text{No}/c=\text{No}) = 1/2$

$P(x_3=\text{Yes}/c=\text{Yes}) = 5/8$; $P(x_3=\text{No}/c=\text{Yes}) = 3/8$

$P(x_3=\text{Yes}/c=\text{No}) = 1/2$; $P(x_3=\text{No}/c=\text{No}) = 1/2$

x_1 \ c	Yes	No
Yes	6/8	1/2
No	2/8	1/2

$P(x_1/c)$

x_2 \ c	Yes	No
Yes	4/8	1/2
No	4/8	1/2

$P(x_2/c)$

x_3 \ c	Yes	No
Yes	5/8	1/2
No	3/8	1/2

* Inference : Calculate Posterior probability for new examples

Probability of infected given contact is 'Yes', symptoms is 'No' and test is 'Yes'

$P(C=\text{Yes}/x_1=\text{Yes}, x_2=\text{No}, x_3=\text{Yes})$

$$= \frac{\overbrace{P(x_1=\text{Yes}, x_2=\text{No}, x_3=\text{Yes} \mid c=\text{Yes})}^{A} \cdot \overbrace{P(c=\text{Yes})}^{B}}{\underbrace{P(x_1=\text{Yes}, x_2=\text{No}, x_3=\text{Yes})}_{C}} \quad \text{—— (1)}$$

$P(C=\text{No}/x_1=\text{Yes}, x_2=\text{No}, x_3=\text{Yes})$ E

$$= \frac{\overbrace{P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes / C = No)}^{D} \cdot \overbrace{P(C = No)}}{\underbrace{P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes)}_{C}} \quad \text{—— (2)}$$

(A) $= P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes \mid C = Yes)$

$= P(x_1 = Yes / C = Yes) \cdot P(x_2 = No / C = Yes) \cdot P(x_3 = Yes / C = Yes)$

[$\because x_1, x_2,$ & x_3 are independent (Naive Assumption)]

$= (6/8) \cdot (4/8) \cdot (5/8)$

$= 15/64$

(B) $P(C = Yes) = 8/10$

(D) $P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes \mid C = No)$

$= P(x_1 = Yes / C = No) \cdot P(x_2 = No / C = No) \cdot P(x_3 = Yes / C = No)$

$= (1/2)(1/2)(1/2)$

$= (1/8)$

(E) $P(C = No) = 2/10$

(C) $P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes)$

$= P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes,\ C = Yes)$

$\quad + P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes,\ C = No)$

[x_1, x_2 & x_3 are conditionally dependent of C; C is marginalized by considering both possible outcomes]

$$\overbrace{P(x_1 = Yes,\ x_2 = No,\ x_3 = Yes / C = Yes)}^{A} \overbrace{P(C = Yes)}^{B}$$

$= P(x_1 = Yes, x_2 = No, x_3 = Yes / C = Yes)\, P(C = Yes)$

$+ P(x_1 = Yes, x_2 = No, x_3 = Yes / C = No)\, P(C = No)$

D E

$$= \left(\frac{15}{64}\right)\left(\frac{8}{10}\right) + \left(\frac{1}{8}\right)\left(\frac{2}{10}\right) = \frac{15}{80} + \frac{2}{80} = \frac{17}{80}$$

$P(C = Yes / x_1 = Yes, x_2 = No, x_3 = Yes)$ — 1

$$= \frac{A \times B}{C}$$

$$= \frac{(15/64) \times (8/10)}{(17/80)}$$

$$= \frac{15}{80} \times \frac{80}{17} = \frac{15}{17}$$

$P(C = No / x_1 = Yes, x_2 = No, x_3 = Yes)$ — 2

$$= \frac{D \times E}{C}$$

$$= \frac{(1/8)(2/10)}{(17/80)}$$

$$= \frac{2}{17}$$

1 > 2 $\Rightarrow$ More likely to be infected