Naive Bayes

☆ Bayes theorem

Likelihood

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Prior

Posterior

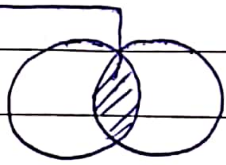
evidence

☆ Proof of Bayes theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad - (1)$$

Now

$$\boxed{A \cap B = B \cap A}$$



$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(B|A) * P(A) = P(B \cap A) = P(A \cap B) \quad - (2)$$

Using (2) in (1)

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

Date __/__/__ Page No.. __

* Naive Bayes Intuition

Dataset

Toss	vs	weather	Result
won	LIV	cloudy	won
won	UTD	sunny	won
lost	BAY	sunny	won
lost	ARE	cloudy	lost

$\underbrace{\text{Toss, vs, weather}}_{X}$

Test = { won, ARE, cloudy }

$$P(W|X) = \frac{P(X|W)\,P(W)}{P(X)}$$

win

$$P(L|X) = \frac{P(X|L)\,P(L)}{P(X)}$$

Both have this in denominator, so useless

* Mathematics behind

$X = \{x_1, x_2, x_3 \ldots x_n\} \rightarrow n$ features

$C = \{C_1, C_2, C_3 \ldots C_k\}$ — k class classification

Acc to Bayes theorem:

$$P(C|X) = \frac{P(X|C)\,P(C)}{P(X)}$$

$P(X) \rightarrow$ again useless so lets remove it.

$$P(C|X) = P(X|C)\,P(C) \quad \text{--- (1)}$$

Also, we know

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B)\,P(B) = P(A \cap B)$$

we use this logic in (1)

$$P(C|X) = \underbrace{P(X \cap C) = P(X, C)}_{\text{"," and "$\cap$" are same}}$$

Again

$$P(C|X) = P(\overset{A}{\boxed{X_1}}, \overset{B}{\boxed{X_2, X_3 \cdots X_n, C_k}})$$

$$= \underbrace{P(X_1|X_2, X_3 \cdots X_n, C_k)}_{a}\,P(X_2, X_3 \cdots X_n, C_k)$$

$$P(C|X) = a\,\underbrace{P(X_2, X_3 \cdots X_n, C_k)}$$

$$P(\overset{A}{\boxed{X_2}}, \overset{B}{\boxed{X_3 \cdots, C_k}}).$$

$$= a\,\underbrace{P(X_2|X_3 \cdots X_n, C_k)}_{b}\,P(X_3 \cdots X_n, C_k)$$

$= ab\, P(X_3 \cdots X_n, C_k)$

In the end

These can be very small

$= P(X_1 | X_2, X_3 - X_n, C_k)\, P(X_2 | X_3, X_4 \cdots X_n, C_k)$

$\cdots\cdots\cdots P(X_{n-1} | X_n, C_k)\, P(X_n | C_k)\, P(C_k)$ — (2)

All of this is very hectic, so we start making assumptions.

∴ Assumption is conditional Independence

$\boxed{P(A|B, C) = P(A|C)}$

(2) will become

$P(C|X) = P(X_1|C)\, P(X_2|C) \cdots P(X_n|C)\, P(C)$

$$\boxed{P(C_i|X) = P(C) \prod_{i=1}^{n} P(x_i|C_i)}$$

$$\hat{y} = \underset{k \in (1, 2 \cdots k)}{\operatorname{argmax}}\, P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

$\boxed{\text{maximum a postirion}}$ rule

MAP

* Naive Bayes with Numerical Data

Height	weight	Gender
172	150	M
180	170	M
165	140	M
190	200	M
139	100	F
145	120	F

← Data

$\{H = 185, W = 170, g = ?\}$

we assume height is a Gaussian distributed RV

$P(M|x) = P(H=185|M)\, P(W=170|M)\, P(M)$

Now for Gaussian Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2x}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

↳ This gives us the probability

Now we can do the same for weight column

Here we took Gaussian Distribution, but you can take multinomial, Poisson etc also. Just the formula changes.