Date
Page

(4) Naive Bayes classifiers are collection of classification algorithm based on Bayes Theorem When we have large training set available then it is moderate to use. Attributes that describes instances are conditionally independent given classification.

Successful Applications :
- Classifying Text Documents
- Diagnosis

Assume target function $f: X \to V$, where each instance x described by attributes $(a_1, a_2 \ldots a_n)$
Most probable value of $f(x)$ is:

$$V_{map} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2 \ldots a_n)$$

$$= \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)}$$

$$= P(a_1, a_2 \ldots a_n | v_j) P(v_j)$$

Naive Bayes Assumption which gives

$$P(a_1, a_2 \ldots a_n | v_j) = \prod P(a_i | v_j)$$

$$V_{NB} = P(v_j) \prod P(a_i | v_j)$$

Algorithm :

N_Bayes (examples)
For each target value v_j
$\hat{P}(v_j) \leftarrow$ estimate $P(v_j)$
For each attribute value a_i of each attribute a
$\hat{P}(a_i | v_j) \leftarrow$ estimate $P(a_i | v_j)$

Classify New Instance (x)

$$v_{NB} = P(v_j) \prod_i P(a_i|v_j)$$

Example: Consider Play Tennis

$\langle$Outlook = sun, Temp = Cool, Humid = high, Wind = Strong$\rangle$

$$v_{NB} = P(v_j) \prod_i P(a_i|v_j)$$

$P(y)\, P(sun|y)\, P(Cool|y)\, P(high|y)\, P(strong|y) = 0.005$

$P(n)\, P(sun|n)\, P(Cool|n)\, P(high|n)\, P(strong|n) = 0.021$

i) Conditional independence assumption is often violated $P(a_1, a_2 \ldots a_n|v_j) = \prod P(a_i|v_j)$

- It works surprisingly well anyway. Note dont need estimated posteriors $\hat{P}(v_j|x)$ to be correct. need only that.

$$\hat{P}(v_j) \prod \hat{P}(a_i|v_j) = P(v_j) P(a_1 \ldots a_n|v_j)$$

- Naive Bayes posterior often unrealistically close to 1 or 0.

ii) what if none of the training instance with target value v_j have attribute value a_i? then

$$\hat{P}(a_i|v_j) = 0$$

$$\hat{P}(v_j) \prod \hat{P}(a_i|v_j) = 0$$

Bayesian Estimation for $\hat{P}(a_i|v_j)$

$$\hat{P}(a_i|v_j) \leftarrow \frac{n_c + mp}{n+m}$$

where, n is no. of training example for $v = v_j$

n_c " " " examples for $v = v_j$ & $a = a_i$

p is prior estimate for $\hat{P}(a_i|v_j)$

m is weight given to prior