0.1 Naive Bayes Classifier

- *Algorithm:* Naive Bayes Classifier(algo. 1)
- *Input:* The training set $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$; $x_j^{(i)}$ is the j^{th} feature of the i^{th} sample;a_{jl} is the l possible values of the j^{th} feature
- *Complexity:* $\mathcal{O}(nk)$
- *Data structure compatibility:* N/A
- *Common applications:* Artificial intelligence

Problem. Naive Bayes Classifier

Naive Bayesian Classification is a classification method based on Bayesian Theorem and Conditional Independence Assumption.

Description

Bayesian Theorem

Bayes's theorem is stated as[1]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{0.1.1}$$

where A and B are events and $P(B) \neq 0$.

Naive Bayes Classifier

Assume input space $X \in \mathbb{R}^n$ is a n dimension vector and the label set of input $Y = \{c_1, c_2, \cdots, c_K\}$. The training set can be denoted as

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\} \tag{0.1.2}$$

which is generated by $P(X, Y)$ independently.

The goal of Naive Bayes is to learn a joint distribution $P(X, Y)$. Specially, it should learn prior distribution and conditional distribution. The prior distribution is

$$P(Y = c_k), \quad k = 1, 2, \cdots, K \tag{0.1.3}$$

The conditional distribution is

$$P(X = x|Y = c_k) = P(X_1 = x_1, \cdots X_n = x_n|Y = c_k) \tag{0.1.4}$$

Naive Bayes makes a conditional independence assumption, which is

$$P(X = x|Y = c_k) = \prod_{j=1}^{n} P(X_= x_j|Y = c_k) \tag{0.1.5}$$

Naive Bayes will use input x and output the class by the learned largest posterior distribution $P(Y = c_k|X = x)$.

The posterior can be calculated by using Bayesian Theorem and equation. 0.1.5

$$P(Y=c_k|X=x) = \frac{P(X=x|Y=c_k)P(Y=c_k)}{\sum_k P(X=x|Y=c_k)P(Y=c_k)} \tag{0.1.6}$$

$$= \frac{P(Y=c_k)\prod_{j=1}^{n} P(X_=x_j|Y=c_k)}{\sum_k P(Y=c_k)\prod_{j=1}^{n} P(X_=x_j|Y=c_k)}, \quad k=1,2,\cdots,K \tag{0.1.7}$$

Then, the Naive Bayes can be represented as

$$y = \arg\max_{c_k} \frac{P(Y=c_k)\prod_{j=1}^{n} P(X_=x_j|Y=c_k)}{\sum_k P(Y=c_k)\prod_{j=1}^{n} P(X_=x_j|Y=c_k)} \tag{0.1.8}$$

$$= \arg\max_{c_k} P(Y=c_k)\prod_{j=1}^{n} P(X_=x_j|Y=c_k) \tag{0.1.9}$$

Maximum Likelihood Estimation(MLE)

We can use MLE to estimate prior probability $P(Y=c_k)$.

$$P(Y=c_k) = \frac{\sum_{i=1}^{N} I(y_i=c_k)}{N}, \quad k=1,2,\cdots,K \tag{0.1.10}$$

Assume possible value set of the j^{th} feature x_j is $\{a_{j1}, a_{j2}, \cdots, a_{jS_j}\}$. The estimation of conditional probability $P(X_j=a_{jl}|Y=c_k)$ is

$$P(X_j=a_{jl}|Y=c_k) = \frac{\sum_{i=1}^{N} I(x_j^{(i)}=a_{jl}, y_i=c_k)}{\sum_{i=1}^{N} I(y_i=c_k)} \tag{0.1.11}$$

$$j=1,2,\cdots,n; \quad l=1,2,\cdots,S_j; \quad k=1,2,\cdots,K \tag{0.1.12}$$

where $x_j^{(i)}$ is the j^{th} feature of the i^{th} sample; a_{jl} is the l possible values of the j^{th} feature.

Algorithm 1: Naive Bayes Classifier

Input : The training set $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$; $x_j^{(i)}$ is the j^{th} feature of the i^{th} sample;a_{jl} is the l possible values of the j^{th} feature

Output: class of the sample x: y

1 Calculate prior distribution by equation. 0.1.10 and conditional distribution by equation. 0.1.12.
2 Calculate posterior distribution with sample $x=(x_1,x_2,\cdots,x_n)^\mathsf{T}$ and equation. 0.1.7.
3 Find the class y of x with equation. 0.1.9.
4 **return** y

References.

[1] M. G. Kendall, A. Stuart, and J. K. Ord. *Kendall's Advanced Theory of Statistics*. USA: Oxford University Press, Inc., 1987. ISBN: 0195205618 (cit. on p. 1).