Simple Bayesian Classifier

Prof.Dr. Songül Varlı
Yıldız Teknik Üniversitesi
Bilgisayar Müh. Bölümü

The Bayesian Theorem

- Let X be a data sample whose class label is unknown.
- Let H be some hypothesis: such that the data sample X belongs to a specific class C.
- We want to determine P(H/X), the probability that the hypothesis H holds given the observed data sample X.
- P(H/X) is the posterior probability representing our confidence in the hypothesis after X is given.

The Bayesian Theorem

- In contrast, P(H) is the prior probability of H for any sample, regardless of how the data in the sample looks.
- The posterior probability P(H|X) is based on more information then the prior probability P(H).
- The Bayesian Theorem provides a way of calculating the posterior probability P(H|X) using probabilities P(H), P(X), and P(X|H).
- The basic relation is

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$

The Bayesian Theorem

- Suppose now that there are a set of m samples S = {S1, S2, …, Sm} (the training data set) where every sample Si is represented as an n-dimensional vector {x1, x2, …, xn}.
- Values xi correspond to attributes A1, A2, …, An, respectively.
- Also, there are k classes C1, C2, …, Ck, and every sample belongs to one of these classes.
- Given an additional data sample X (its class is unknown), it is possible to predict the class for X using the highest conditional probability P(Ci|X), where i = 1. …, k.

The Bayesian Theorem

- That is the basic idea of **Naïve-Bayesian** Classifier. These probabilities are computed using Bayes Theorem:

$$P(Ci \mid X) = \frac{P(X \mid Ci)P(Ci)}{P(X)}$$

- As P(X) is constant for all classes, only the product P(X|Ci) · P(Ci) needs to be maximized. We compute the prior probabilities of the class as
- P(Ci) = number of training samples of class Ci/m (m is total number of training samples).

The Bayesian Theorem

- Because the computation of P(X|Ci) is extremely complex, especially for large data sets, the Naïve assumption of conditional independence between attributes is made.
- Using this assumption, we can express P(X|Ci) as a product:

$$P(X \mid Ci) = \prod_{t=1}^{n} P(X_t \mid Ci)$$

- where xi are values for attributes in the sample X. The probabilities $P(X_t|Ci)$ can be estimated from the training data set.

Example Dataset for Naive Bayes Classifier

Table 10.4 • **Data for Bayes Classifier**

Magazine Promotion	Watch Promotion	Life Insurance Promotion	Credit Card Insurance	Sex
Yes	No	No	No	Male
Yes	Yes	Yes	Yes	Female
No	No	No	No	Male
Yes	Yes	Yes	Yes	Male
Yes	No	Yes	No	Female
No	No	No	No	Female
Yes	Yes	Yes	Yes	Male
No	No	No	No	Male
Yes	No	No	No	Male
Yes	Yes	Yes	No	Female

- Consider the following new sample to be classified:
  - Magazine Promotion = *Yes*
  - Watch Promotion = *Yes*
  - Life Insurance Promotion = *No*
  - Credit Card Insurance = *No*
  - Sex = ?

We have two hypothesis to be tested.

- One hypothesis states the credit card holder is male
- The second hypothesis sees the sample as a female card holder

Table 10.5 • **Counts and Probabilities for Attribute Sex**

	Magazine Promotion		Watch Promotion		Life Insurance Promotion		Credit Card Insurance	
Sex	Male	Female	Male	Female	Male	Female	Male	Female
Yes	4	3	2	2	2	3	2	1
No	2	1	4	2	4	1	4	3
Ratio: yes/total	4/6	3/4	2/6	2/4	2/6	3/4	2/6	1/4
Ratio: no/total	2/6	1/4	4/6	2/4	4/6	1/4	4/6	3/4

- In oreder to determine which hypothesis is correct, we apply Bayes classifier to compute a probability for each hypothesis.

$$P(sex = female \mid E) = \frac{P(E \mid sex = female)\, P(sex = female)}{P(E)}$$

$$P(sex = male \mid E) = \frac{P(E \mid sex = male)\, P(sex = male)}{P(E)}$$

P(sex=female|X)=?

$$P(sex = female \mid E) = \frac{P(E \mid sex = female)\, P(sex = female)}{P(E)}$$

- $P(magazine\ promotion = yes \mid sex = female) = 3/4$
- $P(watch\ promotion = yes \mid sex = female) = 2/4$
- $P(life\ insurance\ promotion = no \mid sex = female) = 1/4$
- $P(credit\ card\ insurance = no \mid sex = female) = 3/4$

- $P(E \mid sex = female) = (3/4)\ (2/4)\ (1/4)\ (3/4) = 9/128$

$P(sex = female \mid E) \approx (9/128)\ (4/10)\ /\ P(E)$

$P(sex = female \mid E) \approx 0{,}0281\ /\ P(E)$

P(sex=male|X)=?

$$P(sex = male \mid E) = \frac{P(E \mid sex = male)\, P(sex = male)}{P(E)}$$

- P(*magazine promotion* = *yes* | *sex* = *male*) = 4/6
- P(*watch promotion* = *yes* | *sex* = *male*) = 2/6
- P(*life insurance promotion* = *no* | *sex* = *male*) = 4/6
- P(*credit card insurance* = *no* | *sex* = *male*) = 4/6

- *P(E* | *sex =male)* = (4/6) (2/6) (4/6) (4/6) = 8/81

***P(sex* = *male* | *E*) ≈ (8/81) (6/10) / *P*(*E*)**

***P(sex* = *male* | *E*) ≈ 0,0593 / *P*(*E*)**

- $P(sex = male \mid E) \approx 0.0593 \,/\, P(E)$
- $P(sex = female \mid E) \approx 0.0281 \,/\, P(E)$

- Because 0.0593>0.0281, Bayes classifier tells us the sample is most likely a male credit card customer.