Machine Learning

Naive Bayes Classifier

Naive Bayes Classification

greatlearning Learning for Life

- Will my flight be on time? It is Sunny, Hot, Normal Humidity, and not Windy!
- Data from the last several times we took this flight

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	Flight On Time
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	Yes
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	No
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No

Probability Review

- If A is any event, then the complement of A, denoted by $\bar{A}$, is the event that A does not occur.
- The probability of A is represented by P(A), and the probability of its complement $P(\bar{A}) = 1 - P(A)$.

- Let A and B be any events with probabilities P(A) and P(B).
  - If you are told that B has occurred, then the probability of A might change. The new probability of A is called the conditional probability of A given B.
  - Conditional probability: $P(A|B) = P(A \text{ and } B) / P(B)$
  - Multiplication rule: $P(A \text{ and } B) = P(A|B)\, P(B)$

Probabilistic Independence

- Probabilistic independence means that knowledge of one event is of no value when assessing the probability of the other.
- The main advantage to knowing that two events are independent is that in that case the multiplication rule simplifies to: P(A and B) = P(A) P(B).

Bayes' Rule

- P(A|B), reads "A given B," represents the probability of A if B was known to have occurred.
- In many situations we would like to understand the relation between P(A|B) and P(B|A).
- You are planning an outdoor event tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. Historically it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. What is the probability that it will rain tomorrow?

use Bayes' Rule:

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

Bayes' Rule Continued

- Let A_1 through A_n be a set of mutually exclusive outcomes.
- The probabilities of the As are $P(A_1)$ through $P(A_n)$. These are called prior probabilities.
- Because an information outcome might influence our thinking about the probabilities of any Ai, we need to find the conditional probability P(Ai|B) for each outcome Ai. This is called the posterior probability of Ai.
- Using Bayes' Rule:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)}$$

Bayes' Rule Continued

- In words, Bayes' rule says that the posterior is the likelihood times the prior, divided by a sum of likelihoods times priors.
- The denominator in Bayes' rule is the probability P(B).

$$\boxed{\text{posterior probability} = \frac{\text{conditional probability} \cdot \text{prior probability}}{\text{evidence}}}$$

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

So will our flight be on time?

greatlearning
Learning for Life

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

On Time?		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Naïve Bayes Classifiers

- Probabilistic models based on Bayes' theorem.
- It is called "naive" due to the assumption that the features in the dataset are mutually independent
- In real world, the independence assumption is often violated, but naïve Bayes classifiers still tend to perform very well
- Idea is to factor all available evidence in form of predictors into the naïve Bayes rule to obtain more accurate probability for class prediction
- It estimates conditional probability which is the probability that something will happen, given that something else has already occurred. For e.g. the given mail is likely a spam given appearance of words such as "prize"
- fBeing relatively robust, easy to implement, fast, and accurate, naive Bayes classifiers are used in many different fields

Naïve Bayes Classifiers - Pros and Cons

- Advantages
  - Simple, Fast in processing and effective
  - Does well with noisy data and missing data
  - Requires few examples for training (assuming the data set is a true representative of the population)
  - Easy to obtain estimated probability for a prediction
- Dis-advantages
  - Relies on and often incorrect assumption of independent features
  - Not ideal for data sets with large number of numerical attributes
  - Estimated probabilities are less reliable in practice than predicted classes
  - If rare events are not captured in the training set but appears in the test set the probability calculation will be incorrect

Gaussian Naive Bayes classifier

- When some of our independent variables are continuous we cannot calculate conditional probabilities!
- In Gaussian Naive Bayes, continuous values associated with each feature (or independent variable) are assumed to be distributed according to a Gaussian distribution
- All we would have to do is estimate the mean and standard deviation of the continuous variable.