Naive Bayes Classifier

Petra Kralj Novak*

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. It assumes the conditional idependence of attribute values given the class:

$$P(v_1, v_2, ..., v_n|c_i) = \prod_{i=1}^{n} P(v_i|c_i)$$

Naive Bayes formula

$$P(c_i|v_1, v_2, ..., v_n) \propto P(c_i) \prod_{j=1}^{n} P(v_j|c_i)$$

Legend:

$c_1, c_2, ..., c_k$	classes
$P(c_1), P(c_2), ..., P(c_k)$	prior probabilities of classes
$v_1, v_1, ..., v_n$	attribute values
$\propto$	is proportional to
$\prod$	product

Classifying a new instance $(v_1, v_2, ..., v_n)$

Let's say that our dataset has k classes $(c_1, c_2, ..., c_k)$ (target variable with k values). The Naive Bayes classifier calculates for each class c_i the conditional probability[1] of class c_i given evidence $(v_1, v_2, ..., v_n)$

$$P(c_i|v_1, v_2, ..., v_n)$$

according to the naive Bayes formula. It classifies the example into the class with the highest value:

$$\text{classification} = \text{argmax}_{c_i} P(c_i) \prod_{j=1}^{n} P(v_j|c_i)$$

*Petra.Kralj.Novak@ijs.si, http://kt.ijs.si/petra_kralj/dmkd.html

[1]a value proportional to the conditional probability

Example

Will the spider catch an ant?

Past experiences of the spider catching ants:

Color	Size	Time	Caught
black	large	day	**YES**
white	small	night	**YES**
black	small	day	**YES**
red	large	night	**NO**
black	large	night	**NO**
white	large	night	**NO**

Ant 1: Color = white, Time = night

$$v_1 = \text{“}Color = white\text{”} = \text{“}C = w''$$
$$v_2 = \text{“}Time = night\text{”} = \text{“}T = n''$$
$$c_1 = YES$$
$$c_2 = NO$$

$$\begin{aligned} P(C_1|v_1, v_2) = & \\ &= P(\text{YES}|C = w, T = n) \\ &= P(\text{YES}) \cdot P(C = w|\text{YES}) \cdot P(T = n|\text{YES}) \\ &= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \\ &= \frac{1}{18} \end{aligned}$$

$$\begin{aligned} P(C_2|v_1, v_2) = & \\ &= P(\text{NO}|C = w, T = n) \\ &= P(\text{NO}) \cdot P(C = w|\text{NO}) \cdot P(T = n|\text{NO}) \\ &= \frac{1}{2} \cdot \frac{1}{3} \cdot 1 \\ &= \frac{1}{6} \end{aligned}$$

The spider will not catch the white ant at night because P(NO| Color = white, Time = night) > P(YES | Color = white, Time = night).

Ant 2: Color = black, Size = large, Time = day

$$v_1 = \text{“}Color = black\text{”} = \text{“}C = b\text{”}$$

$$v_2 = \text{“}Size = large\text{”} = \text{“}S = l\text{”}$$

$$v_3 = \text{“}Time = day\text{”} = \text{“}T = d\text{”}$$

$$c_1 = YES$$

$$c_2 = NO$$

$$\begin{aligned}
P(C_1|v_1, v_2, v_3) = & \\
&= P(\text{YES}|C = b, S = l, T = d) \\
&= P(\text{YES}) \cdot P(C = b|\text{YES}) \cdot P(S = l|\text{YES}) \cdot P(T = d|\text{YES}) \\
&= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} \\
&= \frac{4}{54} = \frac{2}{27}
\end{aligned}$$

$$\begin{aligned}
P(C_2|v_1, v_2, v_3) = & \\
&= P(\text{NO}|C = b, S = l, T = d) \\
&= P(\text{NO}) \cdot P(C = b|\text{NO}) \cdot P(S = l|\text{NO}) \cdot P(T = d|\text{NO}) \\
&= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{3} \cdot 0 \\
&= 0
\end{aligned}$$

The spider will catch the large black ant at night because p(Caught=YES | Color = black, Size = large, Time = day) > p(Caught=NO | Color = black, Size = large, Time = day).

Probability estimates

The simplest way of estimating the probability of an event e from data is to count the number of times an event occurred ($|e|$) and divide by the total number of trials n. This is referred to as the *relative frequency* of the event.

$$\text{relative frequency} = \frac{|e|}{n}$$

Such estimates are unreliable when estimated on small samples. For example, in our spider case, $P(\text{Time} = \text{day}|\text{Class} = \text{NO}) = 0/3 = 0$. Is it really impossible that a spider catches an ant during the day? Or is just that the sample to small? In our the Naive Bayes case, even if all other evidence was in favor of catching the ant, if it was during the day, the classifier would consider this impossible.

We can overcome this by using the Laplace probability estimate. Laplace estimate assumes equal prior probability of events if no evidence is given. The more evidence it has, the more it is close to relative frequency.

$$\text{Laplace estimate} = \frac{|e| + 1}{n + k}$$

, where k is the number of all possible outcomes.

In our spider example $P(\text{Time} = \text{day}|\text{Class} = \text{NO})$, there are three trials with Class = NO ($n = 3$), there are 0 cases where Time = day and Class = NO ($|e| = 0$) and there are two possible values of the attribute Time: day and night ($k = 2$).

$$P(\text{Time} = \text{day}|\text{Class} = \text{NO}) = \frac{0 + 1}{3 + 2} = \frac{1}{5}$$

If there were 300 cases when the spider didnt catch ants at night:

$$P(\text{Time=day}|\text{Class=NO}) = (0 + 1)/(300 + 2) = 1/302 = 0.003$$

With Laplace estimate, probabilities can never be 0.