Naive Bayes classifier

Tuan Nguyen

April 20, 2023

Overview

Classification problem

Bayes theorem

Naive Bayes algorithm

Relevant Issues

Classification problem

The goal in classification is to take an input vector x with d features $x = [x_1, x_2, ..., x_d]^T$ and to assign it to one of K discrete classes C_k where k = 1,...,K.

We need to calculate the probability of a given sample belonging to each class

$$p(y = C_k|x) = p(C_k|x)$$

$$\sum_{k=1}^{K} p(C_k|x) = 1$$

Then we pick the class with the highest probability

$$c = \arg\max_{C_i} p(y = C_k|x)$$

Bayes' theorem

$$
\begin{aligned}
c &= \arg\max_{C_k} p(C_k|x) \\
&= \arg\max_{C_k} \frac{p(x|C_k)p(C_k)}{p(x)} \\
&= \arg\max_{C_k} p(x|C_k)p(C_k) \\
&= \arg\max_{C_k} \prod_{i=1}^{d} p(x_i|C_k)p(C_k)
\end{aligned}
$$

Assumption: Features in the sample are independent!!!

In the training phase, based on the dataset, we will calculate all components, $p(C_k), p(x_i|C_k)$

Training

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Figure 1: Dataset

Training (cont.)

$$p(C_1) = p(Play = Yes) = \frac{number\_of\_samples\_play\_tennis}{total\_number} = \frac{9}{14}$$

$$p(C_0) = p(Play = No) = 1 - p(C_1) = \frac{5}{14}$$

Each component

Outlook	**Play=Yes**	**Play=No**
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Exercise: Calculate the table for each feature Temperature, Humidity, Wind.

Prediction

Given a new instance, x=(Outlook=Sunny, Temperature=Cool, Humidity=High, Wind=Strong). We need to predict whether the player plays tennis or not.

Lookup table

p(Outlook=Sunny|Yes) = 2/9	p(Outlook=Sunny|No) = 3/5
p(Temperature=Cool|Yes) = 3/9	p(Temperature=Cool|No) = 1/5
p(Huminity=High|Yes) = 3/9	p(Huminity=High|No) = 4/5
p(Wind=Strong|Yes) = 3/9	p(Wind=Strong|No) = 3/5
p(Play=Yes) = 9/14	p(Play=No) = 5/14

Then we calculate and compare p(Yes|x) and p(No|x), then make the prediction

Zero conditional probability Problem

- If no example contains the attribute value, the probability will be zero, p(Overcast|No) = 0
- Then during the testing phase, the posterior of the example containing this attribute will be zero, p(No|x[Overcast]) = 0
- Laplace smoothing

$$p(x_i|C_k) = \frac{N_{ik} + \alpha}{N_k + d\alpha}$$

α is a positive number, default value is 1.

- When there are a lot of feature, the multiplication of probability could lead to 0, we can take the log then compare.

$$\log p(C_k|x) = \log p(C_k) + \sum_{i=1}^{d} \log p(x_i|C_k)$$

Continuous-valued Input Attributes

- Numberless values for an attribute
- Conditional probability modeled with the normal distribution

$$p(x_i|C_k) = \frac{1}{\sqrt{2\pi}\sigma_{ik}} \exp\left\{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}\right\}$$

- In the training phase, we need to find σ_{ik}, μ_{ik} for each feature in each category based on the dataset.