INT104 Artificial Intelligence

Naïve Bayes

Fang Kang
Sichen Liu
Shengchen Li

Xi'an Jiaotong-Liverpool University

12th Jan 2022

Aims

After this lecture, you should be able to

- fully understand principles of Naïve Bayes
- classify samples with Naïve Bayes

Bayes' Rule

The famous Bayes' Rule states the relationship between prior probability distribution and posterior probability distribution

$$P(c|\mathbf{x}) = \frac{P(c)P(\mathbf{x}|c)}{P(\mathbf{x})} \tag{1}$$

where c is considered as a class, $\mathbf{x}$ is considered as a set of samples

- $P(c)$ is named as prior probability
- $P(\mathbf{x}|c)$ is named as class-conditional probability (CCP, also known as "likelihood")
- $P(c|\mathbf{x})$ is named as posterior probability
- $P(\mathbf{x})$ is considered as evidence factor (observation)

Bayes' Rule can be used for various purposes such as parameter estimation, classification and model selection.

Bayes' Rule for Classification

How can we make use of Bayes' Rule for Classification?

- We want to maximise the posterior probability of observations
- This method is named MAP estimation (Maximum a posteriori)

The posterior is simply the CCP times the prior and then normalised.

Bayes' Rule for Classification

As we are discussing classification problem, the representation for classification should be presented.

Presume that $x \in D_c$ means that a sample x belongs to class c where all samples belong to class c form dataset D_c

Recall Bayes' Rule

$$PosteriorProbability = \frac{CCP \times PriorProbability}{Observation} \tag{2}$$

As the observation is same (the same training dataset), we have

$$PosteriorProbability \propto CCP \times PriorProbability \tag{3}$$

Bayes' Rule for Classification

So we need to find the CCP and the prior probability

According to Law of Large Numbers, the prior probability can be taken as the probability resulted from the frequency of observations

So we only care about the term

$$p(D|\Theta) = p(c) \prod_{c} \prod_{x_c \in D_c} p(x_c|\Theta_c) \tag{4}$$

Naïve Bayes Classifier

Calculate the term $p(x_c|\Theta_c)$ is never an easy task

A way to simplify the process is to assume the conditions / features of Θ_c are independent to each other

Assume that $\Theta_c = (\theta_{c1}, \theta_{c2}, \ldots, \theta_{cl})$, we have

$$p(x_c|\Theta_c) = \prod_{i=1}^{l} p(c|\theta_i) \tag{5}$$

Example

Naïve Bayes

Outlook	Temprature	Humidity	Windy	Play
Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No
Sunny	Hot	High	True	No
Sunny	Hot	High	False	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Will you play on the day of Mild?

Solution

Naïve Bayes

Temp.	Yes	No	p
Hot	2	2	0.28
Mild	4	2	0.43
Cool	3	1	0.28
p	0.64	0.36	

By this table we have $p(\text{Mild}|\text{Yes}) = \frac{4}{9} = 0.44$ and $p(\text{Mild}|\text{No}) = \frac{2}{5} = 0.4$

$$\text{Posterior } p(\text{Yes}|\text{Mild}) = \frac{p(\text{Mild}|\text{Yes})p(\text{Yes})}{p(\text{Mild})} = \frac{0.44 \times 0.64}{0.43} = 0.65$$

$$\text{Posterior } p(\text{No}|\text{Mild}) = \frac{p(\text{Mild}|\text{No})p(\text{No})}{p(\text{Mild})} = \frac{0.4 \times 0.36}{0.43} = 0.33$$

As $p(\text{Yes}|\text{Mild}) > p(\text{No}|\text{Mild})$, it is likely to play.

Exercise

Will the following condition be considered as a proper day for play?

- `Sunny, Windy`

- `Overcast, Normal Humidity & Cool`

CCP Tables

Outlook	Yes	No	p
Overcast	4	0	0.28
Rainy	3	2	0.36
Sunny	2	3	0.36
p	0.64	0.36	

Wind	Yes	No	p
True	3	3	0.43
False	6	2	0.57
p	0.64	0.36	

Humid	Yes	No	p
Normal	6	1	0.5
High	3	4	0.5
p	0.64	0.36	

Solution

`Sunny, Windy`

$$p(\text{Yes}|\text{Sunny, Windy}) = \frac{p(\text{Sunny, Windy}|\text{Yes})p(\text{Yes})}{p(\text{Sunny, Windy})} \propto$$

$p(\text{Sunny, Windy}|\text{Yes})p(\text{Yes}) = p(\text{Sunny}|\text{Yes})p(\text{Windy}|\text{Yes})p(\text{Yes}) = 0.22 \times 0.33 \times 0.64 = 0.05$

$$p(\text{No}|\text{Sunny, Windy}) = \frac{p(\text{Sunny, Windy}|\text{No})p(\text{No})}{p(\text{Sunny, Windy})} \propto$$

$p(\text{Sunny, Windy}|\text{No})p(\text{No}) = p(\text{Sunny}|\text{No})p(\text{Windy}|\text{No})p(\text{No}) = 0.6 \times 0.6 \times 0.36 = 0.13$

As $p(\text{Yes}|\text{Sunny, Windy}) < p(\text{No}|\text{Sunny, Windy})$, so the combination of weather is unlikely to be suitable for playing

`Rainy, Normal Humidity & Cool`

$p(\text{Yes}|\text{Rainy, Normal, Cool}) = \dfrac{p(\text{Rainy, Normal, Cool}|\text{Yes})p(\text{Yes})}{p(\text{Rainy, Normal, Cool})} \propto$
$p(\text{Rainy, Normal, Cool}|\text{Yes})p(\text{Yes}) =$
$p(\text{Rainy}|\text{Yes})p(\text{Normal}|\text{Yes})p(\text{Cool}|\text{Yes})p(\text{Yes}) = 0.33 \times 0.67 \times 0.33 \times$
$0.64 = 0.047$

$p(\text{No}|\text{Rainy, Normal, Cool}) = \dfrac{p(\text{Rainy, Normal, Cool}|\text{No})p(\text{No})}{p(\text{Rainy, Normal, Cool})} \propto$
$p(\text{Rainy, Normal, Cool}|\text{No})p(\text{No}) =$
$p(\text{Rainy}|\text{No})p(\text{Normal}|\text{No})p(\text{Cool}|\text{No})p(\text{No}) = 0.4 \times 0.2 \times 0.2 \times 0.36 =$
0.0058

As $p(\text{Yes}|\text{Rainy, Normal, Cool}) > p(\text{No}|\text{Rainy, Normal, Cool})$, it is likely to play