Machine Learning

Lecture 12: Naïve Bayes Classifier

COURSE CODE: CSE451

2023

Course Teacher

Dr. Mrinal Kanti Baowaly

Associate Professor

Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh.

Email: mkbaowaly@gmail.com

Definition: Naïve Bayes Classifier

- A probabilistic classifier based on applying Bayes’ theorem
- The reason why it is called ‘Naïve’ because it requires rigid independence assumption between input variables/attributes
- Two specific assumptions are required for the attributes:
  - ✓ Attributes are statistically independent given the class value
  - ✓ Attributes are equally important

Bayes' theorem

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis.

Bayes' theorem is stated mathematically as the following equation:

Likelihood

Prior Probability

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

Evidence Probability

Posterior Probability

where A and B are events and $P(B) \neq 0$.

- $P(A \mid B)$ is a conditional probability: the likelihood of event A occurring given that B is true.
- $P(B \mid A)$ is also a conditional probability: the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other

» Read Conditional Probability

Naïve Bayes Classifier

- Given a problem instance X to predict the class labels Y. In the Bayes' theorem, if the evidence (B) is represented by an instance (X) and the hypothesis (A) is represented by a class label $y \in Y$, then the probability of the class label y given an instance X is:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

- If we have multiple features i.e., $X = (x_1, x_2, x_3, \dots, x_n)$ then the Bayes' theorem can be rewritten as:

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)}$$

How to classify with Naïve Bayes Classifier

- For the classification, Naïve Bayes Classifier finds the probability of all class labels and pick the most probable one to label the instance
- Suppose, we have two class labels, $Y = \{yes, no\}$ and an instance X
- Calculate posterior probabilities: $P(yes|X)$ and $P(no|X)$
- If $P(yes|X)$ > $P(no|X)$, then X is labeled/classified as yes otherwise as no

Example: Classify with Naïve Bayes Classifier

Problem: If the weather is sunny then can players play or not?

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Weather	Play
Sunny	?

Solution: Find P(Yes|Sunny) and P(No|Sunny)

Example: Classify with Naïve Bayes Classifier (Cont..)

- **P(Yes|Sunny) = P(Sunny|Yes) * P(Yes) / P (Sunny)**
  Here, P(Sunny|Yes) = 3/9 = 0.33, P(Yes)= 9/14 = 0.64, P(Sunny) = 5/14 = 0.36
  Now, P(Yes|Sunny) = 0.33 * 0.64 / 0.36 = 0.60
- **P(No|Sunny) = P(Sunny|No) * P(No) / P (Sunny)**
  Here, P(Sunny|No) = 2/5 = 0.40, P(No)= 5/14 = 0.36, P(Sunny) = 5/14 = 0.36
  Now, P(No|Sunny) = 0.40* 0.36/ 0.36 = 0.40

❖ We can see that P(Yes|Sunny) > P(No|Sunny)

❖ So if the weather is sunny then players can play the sport.

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Now it’s your turn

Problem 1: If the weather is overcast then can players play or not?

Problem 2: If the weather is rainy then can players play or not?

Example: Classify with Naïve Bayes Classifier (In case of multiple features)

Suppose we have a Day with the following values :

Outlook = Rain

Humidity = High

Wind = Weak

Play = ?

No need to calculate this probability (Evidence)

Let X = **(**Outlook=Rain, Humidity=High, Wind = Weak**)**

Find, **P(Yes|X)** = P(X|Yes) * P(Yes) / P(X) and **P(No|X)** = P(X|No) * P(No) / P(X)

Now, P(X|Yes) * P(Yes)

= P(Outlook=Rain, Humidity=High, Wind = Weak|Yes) * P(Yes)

= P(Outlook = Rain|Yes)*P(Humidity= High|Yes)* P(Wind= Weak|Yes)*P(Yes)

Solution: [Dzone - Naive Bayes Tutorial]

Estimating conditional probabilities for continuous attributes

- A Gaussian distribution is usually chosen to represent the class conditional probabilities for continuous attributes
- For each class y, the class conditional probability for x_i

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

where μ represents mean and σ^2 represents variance.

HW: Zero frequency problem

- What is Zero frequency problem in Naïve Bayes Classifier?
- How to handle with Zero frequency problem?

Types of Naïve Bayes Classifier

- Multinomial Naive Bayes : When features are discrete count variables / categorical
- Bernoulli Naive Bayes : When feature vectors are binary (i.e. zeros and ones)
- Gaussian Naive Bayes : When features follow a normal distribution

»Read Normal Distribution

Adv. & Disadv. of Naïve Bayes Classifier

Advantage

- Works surprisingly well
- Simple
- Handling missing value is easier
- Robust to irrelevant attributes

Disadvantage

- Can’t handle dependent variables
- Suffers from “Zero Frequency” problem

Some Learning Materials

[Naïve Bayes]

[Naive Bayes Tutorial: Naive Bayes Classifier in Python]

[Naive Bayes Classification using Scikit-learn]