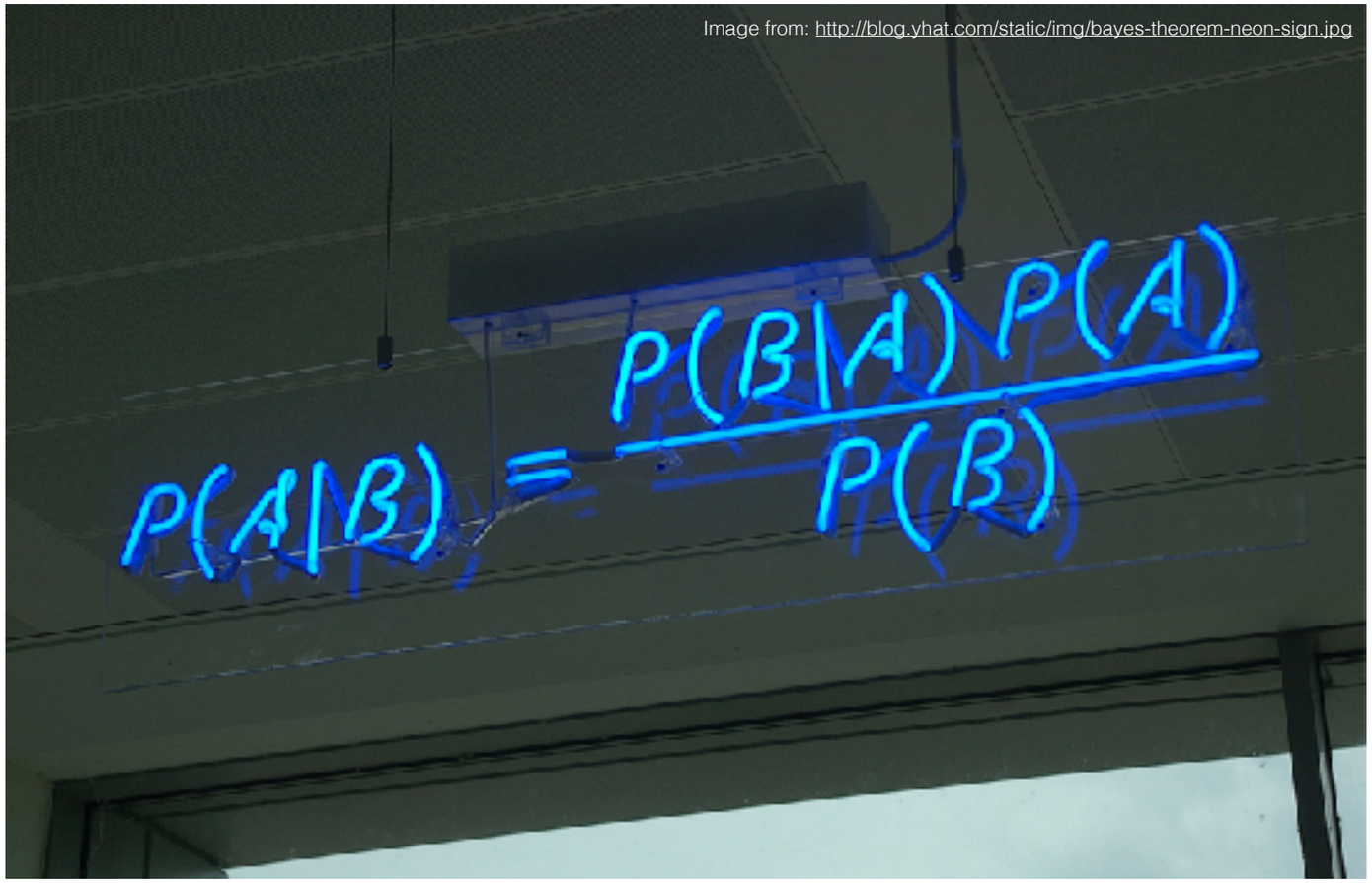
CO3091 - Computational Intelligence and Software Engineering

Lecture 18



Naïve Bayes — Part I

Leandro L. Minku

Announcements

- Three lectures given on Wednesdays from weeks 13-15.
- No lectures from Tuesday to Friday on week 16.
- Schedule goes back to normal now (i.e., no lectures on Wednesdays).
- The number of software engineering problems that can be solved using computational intelligence is ever increasing!
  - Software effort estimation.
  - Estimation of issue resolution time in Agile development.
  - Software defect prediction.
  - Test failure prediction for test case prioritisation.
  - Classifying commits into maintenance activities (corrective, perfective, adaptive).
  - …

Overview

- Machine learning and probabilities
- Bayes theorem
- Naïve bayes:
  - Categorical attributes
- Topic to be continued in the next lecture

Machine Learning and Probabilities

Data are generated from some underlying process.

Images from: http://kaushik88.github.io/assets/img/blog/personal/process.png and https://d1.awsstatic.com/cloud-storage/Deck_Databases.28cb2c2b84fa6f5893cae60ea243c32acf53f366.png

Machine Learning and Probabilities

We can assume that this process generates data based on an unknown probability distribution.

Probability of output class being C given that the input values are **F**

Probability of observing output class C

$$P(\mathbf{F},C) = P(\mathbf{F})\ P(C|\mathbf{F}) = P(C)\ P(\mathbf{F}|C)$$

Probability of observing input values **F** with output class C

Probability of observing input values **F**

Probability of input values being **F** given that the output class is C

Machine Learning and Probabilities

We can assume that this process generates data based on an unknown probability distribution.

$$P(\mathbf{F},C) = P(\mathbf{F})\ P(C|\mathbf{F}) = P(C)\ P(\mathbf{F}|C)$$

Our machine learning task can then be seen as learning probabilities, which can then be used for making predictions.

Learning Probabilities

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Intuitively, if we ignore the attribute Flowers, the probability of Gender = Female when Hair = Long based on the data above should be larger than the probability of Gender = Female when Hair = Short.

We can learn probabilities by keeping track of the frequencies associated to different input and output values.

Illustrative Example for One Input Attribute

Training Set

Person	x_1 (Flowers)	y (Gender)
P1	Likes	Female
P2	Likes	Female
P3	!Like	Female
P4	!Like	Male
P5	!Like	Male

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

Illustrative Example for One Input Attribute

Training Set

Person	x_1 (Flowers)	y (Gender)
P1	Likes	Female
P2	Likes	Female
P3	!Like	Female
P4	!Like	Male
P5	!Like	Male

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

P(Likes|Male) = 0 / 2

P(!Likes|Male) = 2 / 2

P(Likes|Female) = 2 / 3

P(!Likes|Female) = 1 / 3

P(Male) = 2 / 5

P(Female) = 3 / 5

How to Make Predictions? Apply Bayes Theorem

Probability of output class being C given that the input values are F

Probability of observing output class C

$$P(\mathbf{F},C) = P(\mathbf{F})\,P(C|\mathbf{F}) = P(C)\,P(\mathbf{F}|C)$$

Probability of observing input values **F** with output class C

Probability of observing input values **F**

Probability of input values being **F** given that the output class is C

Calculate P(C|**F**) for each class and then predict the class associated to the maximum P(C|**F**).

Example of Applying Bayes Theorem

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

New example: (Likes, y = ?)

$$P(C|\mathbf{F}) = \frac{P(C)\,P(\mathbf{F}|C)}{P(\mathbf{F})}$$

Example of Applying Bayes Theorem

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

New example: (Likes, y = ?)

P(Female | Likes) = ?

P(Male | Likes) = ?

$$P(C|F) = \frac{P(C)\,P(F|C)}{P(F)}$$

Example of Applying Bayes Theorem

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

$$P(C|F) = \frac{P(C)\,P(F|C)}{P(F)}$$

New example: (Likes, y = ?)

$$P(\text{Male} \mid \text{Likes}) = \frac{P(\text{Male})P(\text{Likes} \mid \text{Male})}{P(\text{Likes})}$$

$$= \frac{2/5 * \ 0/2}{2/5}$$

$$= 0\%$$

Example of Applying Bayes Theorem

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

New example: (Likes, y = ?)

P(Female | Likes) =

$$P(C|F) = \frac{P(C)\,P(F|C)}{P(F)}$$

Example of Applying Bayes Theorem

Model

Frequency Table	Gender = Male	Gender = Female	Total:
Flowers = Likes	0	2	2
Flowers = !Like	2	1	3
Total:	2	3	5

$$P(C|F) = \frac{P(C)\,P(F|C)}{P(F)}$$

New example: (Likes, y = ?)

$$P(\text{Female} \mid \text{Likes}) = \frac{P(\text{Female})\;\; P(\text{Likes} \mid \text{Female})}{P(\text{Likes})}$$

$$= \frac{3/5 * \;\; 2/3}{2/5}$$

$$= 100\%$$

Example of Prediction Based on the Bayes Theorem

New example: (Likes, y = ?)

$$P(\text{Male} \mid \text{Likes}) = \frac{P(\text{Male})\, P(\text{Likes} \mid \text{Male})}{P(\text{Likes})}$$

$$= \frac{2/5 * 0/2}{2/5}$$

$$= 0\%$$

$$P(\text{Female} \mid \text{Likes}) = \frac{P(\text{Female})\, P(\text{Likes} \mid \text{Female})}{P(\text{Likes})}$$

$$= \frac{3/5 * 2/3}{2/5}$$

$$= 100\%$$

Predicted class = Female

Simplifying Bayes Theorem for Prediction Purposes

New example: (Likes, y = ?)

$$P(\text{Male} \mid \text{Likes}) = \frac{P(\text{Male})\, P(\text{Likes} \mid \text{Male})}{P(\text{Likes})}$$

$$= \frac{2/5 * 0/2}{2/5}$$

$$= 0\%$$

$$P(\text{Female} \mid \text{Likes}) = \frac{P(\text{Female})\, P(\text{Likes} \mid \text{Female})}{P(\text{Likes})}$$

$$= \frac{3/5 * 2/3}{2/5}$$

$$= 100\%$$

P(F) will be the same for all classes. Given that the relative probabilities are used for prediction, we do not need to compute P(F).

Simplifying Bayes Theorem for Prediction Purposes

New example: (Likes, y = ?)

P(Male | Likes) = P(Male) P(Likes | Male)

= 2/5 * 0/2

= 0%

P(Female | Likes) = P(Female) P(Likes | Female)

= 3/5 * 2/3

= 40%

Predicted class = Female

Bayes Theorem for n Input Attributes, where n ≥ 1

$$P(C|\mathbf{F}) = \frac{P(C)\,P(\mathbf{F}|C)}{P(\mathbf{F})}$$

$$P(C|F_1,\ldots,F_n) = \frac{P(C)\,P(F_1,\ldots,F_n|C)}{P(F_1,\ldots,F_n)}$$

where
P represents a probability calculated based on the frequency tables,
C represents a class,
Fi represents the value of input attribute x_i, i ∈ {1, 2,...,n}, and
n is the number of input attributes.

Example

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Problem: number of possible combinations of input attribute values becomes very large when the number of input attributes and values is large.

Model

Frequency Table	Gender = Female	Gender = Male	Total:
Likes and Long	2	0	2
Likes and Short	0	0	0
!Like and Long	1	0	1
!Like and Short	0	2	2
Total:	3	2	5

Naïve Bayes

- Assumes that each input attribute is conditionally independent of all other input attributes given the output.
- Conditional independence:
  - x_1 is conditionally independent of x_2 given y if, for any value of the input attributes and output, the following is satisfied:
  - P(x1 | x2, y) = P(x1 | y)

If we know the value of y, we don't need to know the value of x_2 in order to determine the value of x_1.

Naïve Bayes

$$P(C|F_1,\ldots,F_n) = P(C)\, P(F_1,\ldots,F_n|C)$$

$$P(C|F_1,\ldots,F_n) = P(C)\, P(F_1|C)\, P(F_2|C)\, \ldots\, P(F_n|C)$$

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

Naïve bayes predicts the class with the maximum $P(C|F_1,\ldots,F_n)$.

Example Using Bayes Theorem

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Problem: number of possible combinations of input attribute values becomes very large when the number of input attributes and values is large.

Model

Frequency Table	Gender = Female	Gender = Male	Total:
Likes and Long	2	0	2
Likes and Short	0	0	0
!Like and Long	1	0	1
!Like and Short	0	2	2
Total:	3	2	5

Example Using Naïve Bayes

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

Number of rows grows linearly with the number of input attribute values.

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2	0	2
!Like	1	2	3
Total:	3	2	5

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3	0	3
Short	0	2	2
Total:	3	2	5

Example of Prediction

- Based on the frequency tables below, determine the predicted class for the following instance:

  (!Like, Long, y = ?)

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

P(Female | !Like,Long) =

= P(Female) P(!Like|Female) P(Long|Female)

= 3/5 * 1/3 * 3/3 = 20%

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2	0	2
!Like	1	2	3
Total:	3	2	5

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3	0	3
Short	0	2	2
Total:	3	2	5

Example of Prediction

- Based on the frequency table in the previous slide, determine the predicted class for the following instance:

  (!Like, Long, y = ?)

$$P(C|F_1,\ldots,F_n) = P(C) \prod_{i=1}^{n} P(F_i|C)$$

P(Male | !Like,Long) =

= P(Male) P(!Like|Male) P(Long|Male)

= 2/5 * 2/2 * 0/2 = 0%

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2	0	2
!Like	1	2	3
Total:	3	2	5

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3	0	3
Short	0	2	2
Total:	3	2	5

Zero Frequency Problem

P(Female | !Like,Long) = P(Female) P(!Like|Female) P(Long|Female)

P(Female | !Like,Long) = 3/5 * 1/3 * 3/3 = 20%

Predicted class = Female

P(Male | !Like,Long) = P(Male) P(!Like|Male) P(Long|Male)

P(Male | !Like,Long) = 2/5 * 2/2 * 0/2 = 0%

Problem: because there were no males with long hair in the training set, no matter the value for Flowers, P(Male | !Like, Long) = 0.

However, clearly males with long hair exist!

Laplace Smoothing

Training Set

Person	x_1 (Flowers)	x_2 (Hair)	y (Gender)
P1	Likes	Long	Female
P2	Likes	Long	Female
P3	!Like	Long	Female
P4	!Like	Short	Male
P5	!Like	Short	Male

To fix this problem, we can add 1 to each of the frequency cells and use that when calculating P(Fi|C).

We calculate P(C) using the original frequencies.

Model

Frequency Table for Flowers	Gender = Female	Gender = Male	Total:
Likes	2+1	0+1	4
!Like	1+1	2+1	5
Total:	5	4	9

Frequency Table for Hair	Gender = Female	Gender = Male	Total:
Long	3+1	0+1	5
Short	0+1	2+1	4
Total:	5	4	9

Further Reading

On the relative value of cross-company and within-company data for defect prediction

Burak Turhan, Tim Menzies, Ayşe B. Bener, Justin Di Stefano

Journal of Empirical Software Engineering

Volume 14 Issue 5, October 2009

Section 3.2 (Naïve Bayes Classifier)

http://readinglists.le.ac.uk/lists/D888DC7C-0042-C4A3-5673-2DF8E4DFE225.html