Georgian

Bayes Theorem and its properties

Garima Malik

February 14, 2022

Outline

Conditional Probability

Bayes Theorem

Classification using Bayes Theorem

Naive Bayes Classifier

Laplace Correction

Conditional Probability

- Conditional probability is the probability of one event occurring with some relationship to one or more other events.
- $P(A|B) = P(AandB)/P(B)$ where P(B) is not equals to 0.
- $P(A|B) = P(A, B)/P(B)$
- $P(A|B) = P(A \bigcap B)/P(B)$

Example

- In a group of 100 sports car buyers, 40 bought alarm systems, 30 purchased bucket seats, and 20 purchased an alarm system and bucket seats. If a car buyer chosen at random bought an alarm system, what is the probability they also bought bucket seats?
- What is the probability a randomly selected person is male, given that they own a pet?

	Have pets	Do not have pets	Total
Male	0.41	0.08	0.49
Female	0.45	0.06	0.51
Total	0.86	0.14	1

Conditional Probability properties

- The joint probability can be calculated using the conditional probability; for example:
  - $P(A, B) = P(A|B) * P(B)$
- This is called the product rule. Importantly, the joint probability is symmetrical, meaning that:
  - $P(A, B) = P(B, A)$
- The conditional probability is not symmetrical; for example:
  - $P(A|B) \neq P(B|A)$

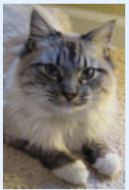
Bayes Theorem

- Bayes Theorem: Principled way of calculating a conditional probability without the joint probability.
  - $P(A|B) = P(B|A) * P(A)/P(B)$
  - $P(B) = P(B|A) * P(A) + P(B|notA) * P(notA)$

Examples

Example: Allergy or Not?

Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:

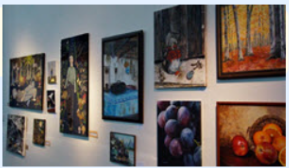


- For people that **really do** have the allergy, the test says "Yes" **80%** of the time
- For people that **do not** have the allergy, the test says "Yes" **10%** of the time ("false positive")

If 1% of the population have the allergy, and **Hunter's test says "Yes"**, what are the chances that Hunter really has the allergy?

Examples

Example: The Art Competition has entries from three painters: Pam, Pia and Pablo



- Pam put in 15 paintings, 4% of her works have won First Prize.
- Pia put in 5 paintings, 6% of her works have won First Prize.
- Pablo put in 10 paintings, 3% of his works have won First Prize.

What is the chance that Pam will win First Prize?

Naming the Terms in the Theorem

- Firstly, in general, the result P(A|B) is referred to as the posterior probability and P(A) is referred to as the prior probability.
- Sometimes P(B|A) is referred to as the likelihood and P(B) is referred to as the evidence.
- This allows Bayes Theorem to be restated as:
  - Posterior = Likelihood * Prior / Evidence

Binary Classification

$$P(C|x) = \frac{P(C)P(x|C)}{P(x)}$$

- $P(C)$: prior
- $P(x|C)$: likelihood
- $P(x)$: evidence
- $P(C|x)$: posterior

$$P(C = 0) + P(C = 1) = 1$$

$$P(x) = P(x|C = 1)P(C = 1) + P(x|C = 0)P(C = 0)$$

$$P(C = 0|x) + P(C = 1|x) = 1$$

Multi Label Classification

$$P(C_i|x) = \frac{P(C_i)P(x|C_i)}{P(x)}$$

$$= \frac{P(C_i)P(x|C_i)}{\sum_{k=1}^{K} P(C_k)P(x|C_k)}$$

$$P(C_i) \geq 0 \text{ and } \sum_{i=1}^{K} P(C_i) = 1$$

$$\text{choose } C_i \text{ if } P(C_i|x) = \max_k P(C_k|x)$$

Prior Probability

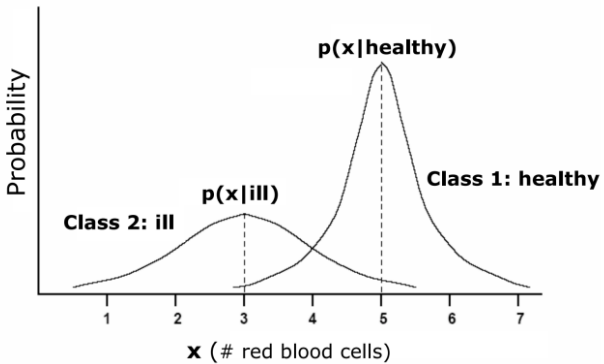
- A prior probability: Knowledge we have about one class before carrying out an experiment.
  - $P(c_i) = \pi_i$ where i = 1 ..... N
- Decision Rule : using only prior probabilities
  - Decide c_1 if $\pi_1 > \pi_2$
  - Decide c_2 if $\pi_2 > \pi_1$
- The probability of having an error in classification is the lower value of π_1 and π_2.

Classification with Prior Probability

- **Classification problem: Discriminate between healthy people and people with anemia.**
- Prior knowledge:
  - 90% of the people is healthy: $\pi_1 = 0.9$
  - 10% of the people is ill: $\pi_1 = 0.1$
- If we have to classify a new patient, which is his/her class?
  - Decide c_1 as $\pi_1 > \pi_2$
- If we have no other information, we have to take this decision. However, nobody will be satisfied if the doctor would decide the state of health without a checkup (or a blood test). It is necessary to use more information:
  - we need measurements relative to the patterns.

Classification with Class Conditional Probability

- Blood test reveals amount of red blood cells.
- The amount of red blood cells is the random variable (x) (We do not have the same number of red blood cells than other people).
- This variable has a Gaussian distribution.



Georgian

Classification with Class Conditional Probability

Classification problem: Discriminate between healthy people and people with anemia.

- Blood test: 4.5 million red blood cells.
- The patient is healthy.
  - $P(x = 4,500,000|c = \text{healthy}) > P(x = 4,500,000|c = \text{ill})$
- If we consider the patient is healthy, the probability he has 4.5 million red blood cells is higher than if we consider he is ill, with the given number of red blood cells

Classification with Class Conditional Probability

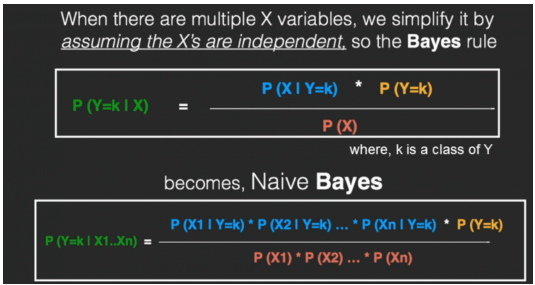
- Bayes Decision Rule :

$$\text{Decide } c_1 \text{ if } P(c_1|x) > P(c_2|x) \left(\text{ or } P(x|c_1)P(x_1) > P(x|c_2)P(x_2) \right)$$

$$\text{Decide } c_2 \text{ if } P(c_2|x) > P(c_1|x) \left(\text{ or } P(x|c_2)P(x_2) > P(x|c_1)P(x_1) \right)$$

Naive Bayes Classifier

- The Bayes Rule provides the formula for the probability of Y given X. But, in real-world problems, you typically have multiple X variables. When the features are independent, we can extend the Bayes Rule to what is called Naive Bayes. It is called 'Naive' because of the naive assumption that the X's are independent of each other. Regardless of its name, it's a powerful formula.



Georgian

Example

- Say you have 1000 fruits which could be either ‘banana’, ‘orange’ or ‘other’. These are the 3 possible classes of the Y variable. We have data for the following X variables, all of which are binary (1 or 0).
  - Long
  - Sweet
  - Yellow

Fruit	Long (x1)	Sweet (x2)	Yellow (x3)
Orange	0	1	0
Banana	1	0	1
Banana	1	1	1
Other	1	1	0
..	..	..	..

Example

- For the sake of computing the probabilities, let's aggregate the training data to form a counts table like this.

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	**Total**
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

Laplace Correction

- The value of P(Orange | Long, Sweet and Yellow) was zero in the above example, because, P(Long | Orange) was zero. That is, there were no ‘Long’ oranges in the training data.
- It makes sense, but when you have a model with many features, the entire probability will become zero because one of the feature’s value was zero. To avoid this, we increase the count of the variable with zero to a small value (usually 1) in the numerator, so that the overall probability doesn’t become zero.
- This correction is called ‘Laplace Correction’. Most Naive Bayes model implementations accept this or an equivalent form of correction as a parameter.

Laplace Correction

- It is also called zero frequency problem
- It is applied on categorical values.
- The generalised formula can be stated as :

$$P(X_i = v_j | C_k) = \frac{n_{ijk} + \lambda}{n_k + \lambda k} \tag{1}$$

- n_ijk is no of examples in C_k where $X_i = v_j$
- n_k is total number of examples in k class
- λ is usually 1
- k is no of classes

Gaussian Naive Bayes

So far we've seen the computations when the X's are categorical. But how to compute the probabilities when X is a continuous variable?

- If we assume that the X follows a particular distribution, then you can plug in the probability density function of that distribution to compute the probability of likelihoods.
- If you assume the X's follow a Normal or Gaussian Distribution, which is fairly common, we substitute the corresponding probability density of a Normal distribution and call it the Gaussian Naive Bayes. You need just the mean and variance of the X to compute this formula.

$$P(X|Y=c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{\frac{-(x-\mu_c)^2}{2\sigma_c^2}}$$

Thank You