Machine Learning

Naive Bayes & Bayes Classifier

Marchelo Bragagnini

cesarbrma91@gmail.com

@MarchBragagnini

Universidad Católica San Pablo

Centro de Investigación e Innovación en Ciencia Computación

Applications

1. Real time prediction
2. Multi class prediction
3. Text classification, spam filtering, sentiment analysis
4. Recommendation systems

Independent Events

Two events are independent if the result of the second event is not affected by the result of the first event. If A and B are independent events, the probability of both events occurring is the product of the probabilities of the individual events.

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

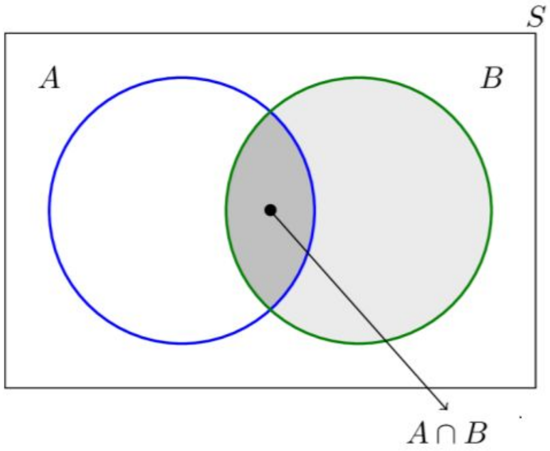
Dependent Events

Two events are dependent if the result of the first event affects the outcome of the second event so that the probability is changed. The probability of both events occurring is the product of the probabilities of the individual events:

$$P(A \text{ and } B) = P(A) \cdot P(B/A)$$

Conditional Probability

The conditional probability of an event A is the probability that the event will occur given the knowledge that an event B has already occurred. This probability is written P(A|B), notation for the probability of A given B.



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

* Jose Castro, Dependent and independent events..

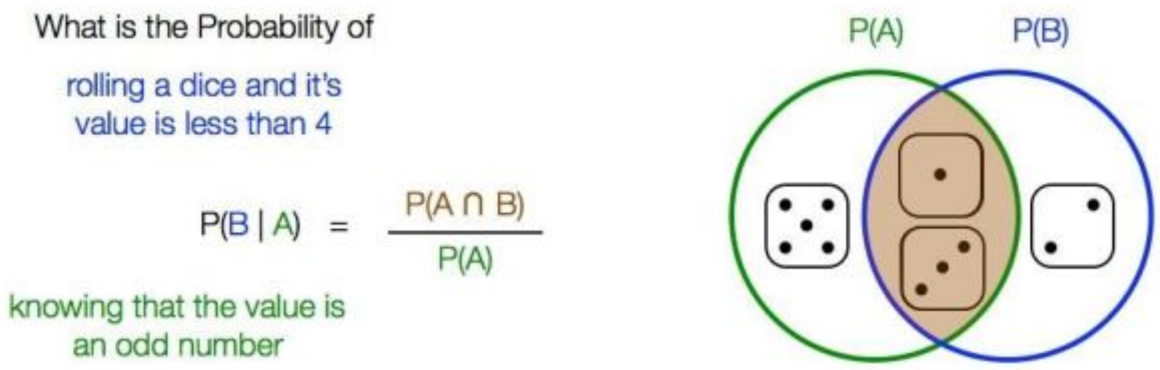
Conditional Probability

Ejercicio: Hallar la probabilidad de que un sólo lanzamiento de un dado resulte en un número menor que 4, (a) no se da ninguna otra información, (b) se da que el lanzamiento resultó en número impar.

* Jose Castro, Dependent and independent events..

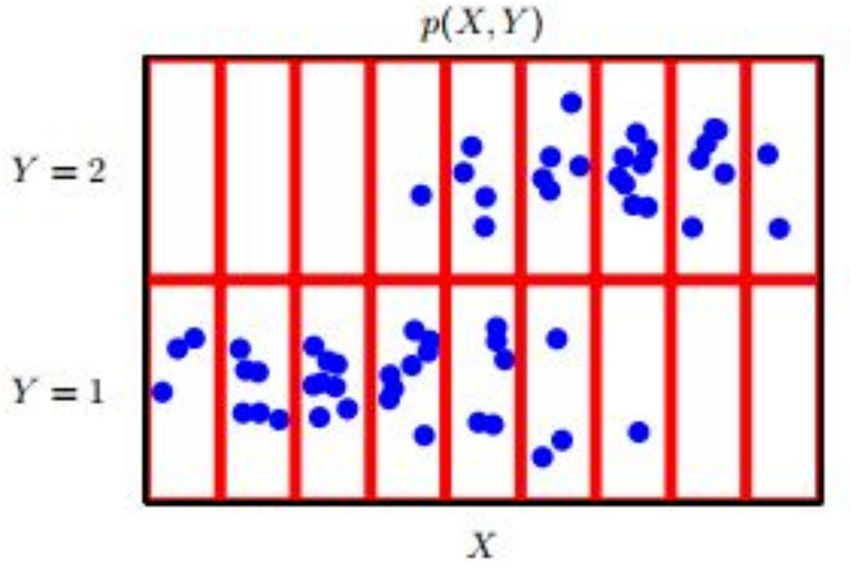
Conditional Probability

Ejercicio: Hallar la probabilidad de que un sólo lanzamiento de un dado resulte en un número menor que 4, (a) no se da ninguna otra información, (b) se da que el lanzamiento resultó en número impar.



* Jose Castro, Dependent and independent events..

Marginal Probability



Marginal Probability

Exercise: An urn contains 5 red balls and 2 green balls. Two balls are drawn one after the other. What is the probability that the second ball is red?

[ ] https://math.mit.edu/~dav/05.dir/class3-prep.pdf

Marginal Probability

Exercise: An urn contains 5 red balls and 2 green balls. Two balls are drawn one after the other. What is the probability that the second ball is red?

Solution:

The sample space is $\Omega = \{rr, rg, gr, gg\}$. Let R1 be the event 'the first ball is red', G1 = 'first ball is green', R2 = 'second ball is red', G2 = 'second ball is green'. We are asked to find P(R2).

$$P(R2) = P(R2|R1)P(R1) + P(R2|G1)P(G1)$$

[ ] **https://math.mit.edu/~dav/05.dir/class3-prep.pdf**

Marginal Probability

Exercise: An urn contains 5 red balls and 2 green balls. Two balls are drawn one after the other. What is the probability that the second ball is red?

Solution:

The sample space is $\Omega = \{rr, rg, gr, gg\}$. Let R1 be the event 'the first ball is red', G1 = 'first ball is green', R2 = 'second ball is red', G2 = 'second ball is green'. We are asked to find P(R2).

$$P(R2) = P(R2|R1)P(R1) + P(R2|G1)P(G1)$$

$$= \frac{4}{6} \cdot \frac{5}{7} + \frac{5}{6} \cdot \frac{2}{7}$$

$$= \frac{30}{42} = \frac{5}{7}.$$

[ ] https://math.mit.edu/~dav/05.dir/class3-prep.pdf

Bayes’ Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

- Comments:
    - Bayes’ rule tells us how to ‘invert’ conditional probabilities, i.e. to find P(B|A) from P(A|B).

Example

Consider a routine screening test for a disease. Suppose the frequency of the disease in the population (base rate) is 0.5%. The test is highly accurate with a 5% false positive rate and a 10% false negative rate.

You take the test and it comes back positive. What is the probability that you have the disease?

[ ] Lesson 3 Christian Córdova
https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading3.pdf

Bayes' Theorem

Example

Events:

D+ = 'you have the disease'

D− = 'you do not have the disease

T+ = 'you tested positive'

T− = 'you tested negative'.

$$P(D+ \mid T+)$$

[ ] Lesson 3 Christian Córdova
https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading3.pdf

Bayes' Theorem

Example

Using :

- P(D+) = 0.005
- P(D-) = _____
- P(T- | D+) = 0.1 (false negative)
- P(T+ | D+) = ___
- P(T+ | D-) = ____ (false positive)

[ ] Lesson 3 Christian Córdova
https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading3.pdf

Bayes' Theorem

Example

$$P(D+ \mid T+) = \frac{P(T+ \mid D+) \cdot P(D+)}{P(T+)}$$

$$P(D+ \mid T+) = \frac{P(T+ \mid D+) \cdot P(D+)}{P(T+ \mid D+) \cdot P(D+) + P(T+ \mid D-) \cdot P(D-)}$$

Naive Bayes

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

Naive Bayes

$$\hat{y} = \underset{k \in \{1,...,m\}}{\operatorname{argmax}} \ p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

Naive Assumption

$$p(C_k \mid x_1, x_2, \dots, x_n) \propto p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

Naive Bayes

Methodology

PARTE 1: Crear el modelo.

Para ello se necesitan **cuatro pasos**:

1. Calcular las probabilidades a priori de cada clase.
2. Para cada clase, realizar un recuento de los valores de atributos que toma cada ejemplo. Se debe distribuir cada clase por separado para mayor comodidad y eficiencia del algoritmo.
3. Aplicar la Corrección de Laplace, para que los valores "cero" no den problemas.
4. Normalizar para obtener un rango de valores [0,1].

PARTE 2:

1. Aplicar la fórmula de Naïve Bayes.

Naive Bayes

Example

Ejemplos	Atr. 1	Atr. 2	Atr. 3	Clase
x1	1	2	1	positiva
x2	2	2	2	positiva
x3	1	1	2	negativa
x4	2	1	2	negativa

For x5 = {1,1,1}, what is the class ?

[ ] http://naivebayes.blogspot.com/

Codes

With Scikit-Learn

1. Gaussian Naive Bayes

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

2. Multinomial Naive Bayes

3. Bernoulli Naive Bayes

$$P(x_i \mid y) = P(i \mid y) x_i + (1 - P(i \mid y))(1 - x_i)$$

With Scikit-Learn

[ ] 1.9. Naive Bayes http://scikit-learn.org/stable/modules/naive_bayes.html

Advantages & When to Use Naive Bayes

Advantages

- They are extremely fast for both training and prediction
- They provide straightforward probabilistic prediction
- They are often very easily interpretable
- They have very few (if any) tunable parameters

Practice

- When the naive assumptions actually match the data (very rare in practice)
- For very well-separated categories, when model complexity is less important
- For very high-dimensional data, when model complexity is less important

[ ] Python Data Science Handbook - Jake VanderPlas

Code

- Naive Bayes with Scikit Learn
  https://github.com/marbramen/100DaysOfMLCode/blob/master/Day%2056%20-%20Naive%20Bayes%20with%20Scikit%20Learn.ipynb
- Naive Bayes with Gaussian Density Function
  https://github.com/marbramen/100DaysOfMLCode/blob/master/Day%2051%20-%20Naive%20Bayes%20with%20Gaussian%20Density%20Function.ipynb

Machine Learning

Naive Bayes & Bayes Classifier

Marchelo Bragagnini

cesarbrma91@gmail.com

@MarchBragagnini

Universidad Católica San Pablo

Centro de Investigación e Innovación en Ciencia Computación