



Naive Bayes Classifier

Outline

- Definition
- Bayes Theorem
- Algorithm
- Event Models
- Example

00101010010101000011110010000100
1000110010000111100101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
1001010010101001010010101010100
10000101010000001001010101010101



Definition

➤ Naive Bayes Classifier

- Probabilistic based classifier with strong naïve (independence) assumptions
- Supervised learning based on *pdf* (probability density function)

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101010010100101010101010
10000101010000000100101010101010



DISCS

Naive

- What do we mean by naïve?
 - Simple but correct
 - Not necessarily efficient

001010100101010000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100010101001001001010
10010100101010010100101010010100
1000010101000000100100100100100



Bayes

➤ Bayes Theorem

- Manipulation of conditional probabilities for interpretation
- Assumes conditionally independent features

001010100101010000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
100101001000010101001001001010
10010100101010010100101010010100
1000010101000000100100100100100



Bayes Theorem

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

001010100101010000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101010010100101010010100
10000101010000001001010101010100



Bayes Theorem

$$Posterior = \frac{(Prior)(Likelihood)}{Evidence}$$

00101010010101000001111010100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101010010100101010101010
100001010100000010010101010101010



Classifier

- Application of Bayes Theorem in a naïve approach to determine the class of a sample data based on prior data
- Classification is based on the higher posterior value.

0010101001010100000111100100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101010010100101010101010
10000101010000001001001001001001



Assumptions

- Features are independent of each other
 - The value of one doesn't affect the value of another
 - Example: Just because you scored good in the midterms doesn't mean you're handsome/pretty.
- Two events are independent if the occurrence of one does not affect the other

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101001010010101010101010
10000101010000000000000000000000



DISCS

Algorithm

➤ Formal notation:

$$P(A \mid B_1 \dots B_n) = \frac{P(A)P(B_1 \dots B_n \mid A)}{P(B_1 \dots B_n)}$$

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
1001010010101001010010101001010
10000101010010010010010010010010



DISCS

Algorithm

- Probability Computation for a class:
 - The share of that class' components against all other classes
 - Normalization factor of that class

Probability Computation for each attribute?

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
0010010101001010010010010010010110
100101001000010101001001001010
1001010010101001010010101001010
1000010101000000100101010101010



Event Models

- What training model will be the Naive Bayes Classifier be based on (event model)?
 - Probability Distributions
 - Multinomial
 - Bernoulli
 - Gaussian
 - Normal
- Fitting data into the proper distribution (Event Model Estimation)

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001001001001110
100101001000010101001001001010
100101001010100101001010101010
10000101010000000000000000000000



Normal/Gaussian

➤ Definition:

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

00101010010101000011110100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
1001010010001010100100101001010
10010100101010010100101010010100
10000101010100000101010101010100



Estimation/Fitting

- Objective: Determine the closest model that your data resembles
- Fitting:
 - Chi Square test
 - Bhattacharyya metric
- Parameters of model will be basis for computing the **posterior probability**

0010101001010100000111100100001100
10001100100001111001101010010101
110010101010100001001100101010100
1001010010010010101010101010101
11100001111010110000000111101001
001001010100101001001010010010110
10010100100001010100100101001010
10010100101010010100101010101010
10000101010000000000000000000000

