{ DEPARTMENT OF
INFORMATION SYSTEMS
AND COMPUTER SCIENCE
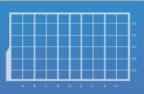
Naive Bayes Classifier

Outline

➢ Definition
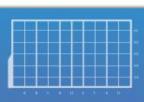
➢ Bayes Theorem

➢ Algorithm

➢ Event Models

➢ Example

Definition

➤ Naive Bayes Classifier

- Probabilistic based classifier with strong naïve (independence) assumptions

- Supervised learning based on *pdf* (probability density function)

Naive

➢ What do we mean by naïve?

    ➢ Simple but correct

    ➢ Not necessarily efficient

Bayes

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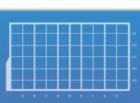 - Assumes conditionally independent features

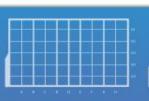# Bayes Theorem

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

Bayes Theorem

$$Posterior = \frac{(\text{Pr}ior)(Likelihood)}{Evidence}$$

Classifier

➢ Application of Bayes Theorem in a naïve approach to determine the class of a sample data based on prior data

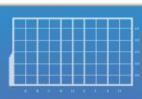➢ Classification is based on the higher posterior value.

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.

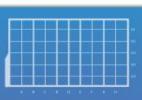➢ Two events are independent if the occurrence of one does not affect the other

Algorithm

➢Formal notation:

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

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?

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)

Normal/Gaussian

➢ Definition:

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

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**