

Naïve Bayes Classification

CS4881 Artificial Intelligence Jay Urbain Credits:

Nazli Goharian, Georgetown; and David Grossman, IIT Machine Learning, Tom Mitchell AIMA, Russell and Norvig

January 27, 201



Naïve Bayes Classifier

- Bayes theorem
- Combines probability of each feature with class label.
- Makes strong independence assumption between features, i.e., features are independent.
- Examples:
 - Determine type of fruit from shape and color
 - Determine credit risk by age, income, and education
 - Determine life style from education and salary Is education and salary really independent?



Naïve Bayes Classifier

- Given a hypothesis, calculate the probability the hypothesis is correct.
- Hypothesis: given x_1 , $x_2 =>$ object is a Peach
- Calculate probability that x_1 , x_2 is a Peach
 - P(H: x₁, x₂ is a Peach)
 - P(H: x₁, x₂ is an Apricot)
 - .
- Calculate each of the probabilities
- 2. Choose the highest probability

January 27, 201



Naïve Bayes Classifier

- P(H/X) Posterior probability of hypothesis H
 - X: {x₁, x_{2...,} x_n} // feature vector
 - Shows the confidence/probability of *H* given *X*
 - x₁: shape=round, x₂: color=orange
 - H: x₁, x₂ is a peach
- P(H) Prior probability of hypothesis H
 - Represents the probability of H just happening, regardless of evidential data.
- E.g. What is the probability of picking a peach from a fruit bin without knowledge of shape and color.

January 27, 2011



Bayes Theorem - Learning

- P(X/H) Likelihood: probability of X conditioned on hypothesis H
 - Shows the confidence/probability of H given X
 - Given H is true (X is a peach) calculate probability that X is round and orange, i.e., x₁=round, x₂=orange.
- P(X) Prior probability of X
 - Represents the probability that sample is round and orange.
 - Use for normalization.

January 27, 201



Bayes Theorem - Classification

Posterior

Probability of X Pri

Prior Probability of class C_i

$$P = (H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Posterio

Probability of class C_i

Prior Probability of X

anuary 27, 2011

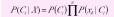


Naïve Bayes Classification

- Hypothesis H is the class C_i.
- Note: P(X) is ignored below as it is constant for all classes.
- Assuming the independence assumption, $P(X/C_i)$ is:

$$P(X \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i)$$

Therefore:



- P(C_i) is the ratio of total samples in class C_i to all samples.
- Divide by P(X) to get correct probabilities does not affect classification.

anuary 27, 2011



Naïve Bayes Classification

- For categorical attribute:
 - $P(x_k/C_i)$ is the frequency of samples having value x_k in class C_k
- For continuous (numeric) attribute:
 - $P(x_k/C_i)$ is calculated via a Gaussian density function.

January 27 201



Naïve Bayes Classification

- Having pre-calculated all $P(x_k/C_i)$, an unknown example X is classified as follows:
 - For all possible classes calculate $P(C_i|X)$
 - 2. Assign X to the class with the highest $P(C_i|X)$

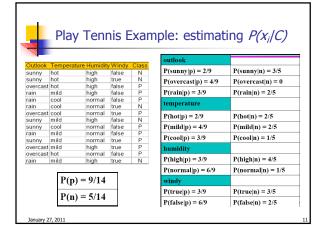
January 27, 2011



Play Tennis?

| Outlook | Temperature | Humidity | Windy | Class |
|----------|-------------|----------|-------|-------|
| sunny | hot | high | false | N |
| sunny | hot | high | true | N |
| overcast | hot | high | false | Р |
| rain | mild | high | false | P |
| rain | cool | normal | false | P |
| rain | cool | normal | true | N |
| overcast | cool | normal | true | Р |
| sunny | mild | high | false | N |
| sunny | cool | normal | false | P |
| rain | mild | normal | false | Р |
| sunny | mild | normal | true | Р |
| overcast | mild | high | true | Р |
| overcast | hot | normal | false | Р |
| rain | mild | high | true | N |

January 27, 201





Play Tennis Example: estimating $P(C_i|x_i)$

- New incoming sample: X = <sunny, cool, high, true>
- $P(\text{play}|X) = P(X|p)*P(p) = \\ P(p)*P(\text{sunny}|p)*P(\text{cool}|p)*P(\text{high}|p)*p(\text{true}|p) \\ 9/14*2/9*3/9*3/9*3/9=0.0053$
- P(no play|X) = P(X|p)*P(n) = P(n)*P(sunny|n)*P(cool|n)*P(high|n)*p(true|n) 5/14 * 3/5 * 1/5 * 4/5 * 3/5 = 0.0206

Class n (no play) has higher probability than class ρ (play) for example X.

Note:

 $P(X) = P(X|no\ play) + P(X|play)$

anuary 27, 2011