Naïve Bayes Classification

Naïve Bayesian Classifier

- Based on Bayes Theorem.
- Combines the impact/probability of each feature on the class label.
- Assumes the independence between the features.
  - » Shape and color of a fruit determining the fruit
  - » Education and salary determining the life style (independence??)

Naïve Bayesian Classifier

Illinois Institute of Technology

- Given a hypothesis, calculating the probability of correctness of that hypothesis.
- Hypothesis: x_1 , x_2 is a Peach.
  - » Calculate the probability that x_1 , x_2 is a Peach.

    P(H: x_1 , x_2 is a Peach)

    P(H: x_1 , x_2 is an Apricot)

    ....

  1. Calculate each of these probabilities.
  2. Choose the highest probability.

Bayes Theorem

- P(H|X) Posterior Probability of hypothesis H
  - » X : x_1 , x_2 , …, x_n
  - » Shows the confidence/probability that suppose X, then the hypothesis is true.
    - – x_1 : shape = round, x_2 : color = orange
    - – H: x_1 , x_2 is a Peach.

- P(H) Prior Probability of hypothesis H
  - » Probability that regardless of data the hypothesis is true.
    - – Regardless of color and shape, it is a Peach.

Bayes Theorem

Illinois Institute of Technology

- P(X|H) Posterior Probability of X conditioned on hypothesis H
  - » Given H is true (X is a Peach), calculate probability that X is round and orange.
- P(X) Prior Probability of X
  - » Probability that sample is round and orange.

Bayes Theorem

$$\underbrace{P=(H|X)}_{\text{Posterior Probability of class } C_i} = \frac{\overbrace{P(X|H)}^{\text{Posterior Probability of X}}\;\overbrace{P(H)}^{\text{Prior Probability of class } C_i}}{\underbrace{P(X)}_{\text{Prior Probability of X}}}$$

Naïve Bayesian Classifier

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

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

- Thus:

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

- $P(C_i)$ is the ratio of total samples in class C_i to all samples.

Naïve Bayesian Classifier

- For Categorical attribute:

  $P(x_k|C_i)$ is the frequency of samples having value x_k in class C_i.

- For Continuous (numeric) attribute:

  $P(x_k|C_i)$ is calculated via a Gaussian density function.

Naïve Bayesian Classifier

- Having pre-calculated all $P(x_k | C_i)$, to classify an unknown sample X:
  - » Step 1: For all classes calculate $P(C_i | X)$.
  - » Step 2: Assign sample X to the class with the highest $P(C_i | X)$.

Play-tennis example: estimating $P(x_i | C)$
(Example from: Tom Mitchell “Machine Learning”)

Illinois Institute of Technology

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

P(p) = 9/14	
P(n) = 5/14	

outlook	
P(sunny|p) = 2/9	**P(sunny|n) = 3/5**
P(overcast|p) = 4/9	**P(overcast|n) = 0**
P(rain|p) = 3/9	**P(rain|n) = 2/5**
temperature	
P(hot|p) = 2/9	**P(hot|n) = 2/5**
P(mild|p) = 4/9	**P(mild|n) = 2/5**
P(cool|p) = 3/9	**P(cool|n) = 1/5**
humidity	
P(high|p) = 3/9	**P(high|n) = 4/5**
P(normal|p) = 6/9	**P(normal|n) = 1/5**
windy	
P(true|p) = 3/9	**P(true|n) = 3/5**
P(false|p) = 6/9	**P(false|n) = 2/5**

Play-tennis example: estimating $P(C_i \mid X)$
(Example from: Tom Mitchell "Machine Learning")

Illinois Institute of Technology

- An incoming sample: X = <sunny, cool, high, true>
- P(play|X) = P(X|p) · P(p) =
  P(p) · P(sunny|p) · P(cool|p) · P(high|p) · P(true|p)=
  9/14 · 2/9 · 3/9 · 3/9 · 3/9 = .0053
- P(Don't play |X) = (X|n) · P(n) =
  P(p) · P(sunny|n) · P(cool|n) · P(high|n) · P(true|n)=
  5/14 · 3/5 · 1/5 · 4/5 · 3/5 = .0206
- Class *n* (don't play) has higer probability than class *p* (play) for sample X.