Naïve Bayes Classification

CS4881 Artificial Intelligence

Jay Urbain, PhD

Credits:

Machine Learning, Tom Mitchell

Nazli Goharian, Georgetown; David Grossman, IIT

Naïve Bayes Classifier

- Bayes theorem
- Combines probability of each feature wrt class label.
- Makes strong independence assumption between features, i.e., independence between features
  - Classify email as spam based on sender, and text
  - Diagnose meningitis based on chest-xray, symptom
  - Classify fruit from shape and color
  - Determine life style from education and salary

Review conditional probability

- Can factor joint probability using the chain rule:
  $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

- And express the joint probability by conditioning on a or b:
  $P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

- … and derive *Bayes* Theorem:
  $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$
  $P(b \mid a) = P(a \mid b)\, P(b) / P(a)$

Naïve Bayes Classifier

- Lets say we have a hyothesis, & we want to calculate the probability of the hypothesis being correct.
- Hypothesis: given feature x_1, x_2 => object is a peach
- Calculate 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

Naïve Bayes Classifier

- P(H|X) *Posterior* probability of hypothesis H
  - X: $\{x_1, x_{2\ldots,} x_n\}$
  - Shows the confidence/probability of H given X
  - x_1: shape=round, x_2: color=orange
  - H: x_1, x_2 is a peach

- P(H) Prior probability of hypothesis H
  - Represents the probability of H just happening, regardless of data.
  - E.g. What is the probability of picking a peach from a fruit bin without knowledge of shape and color.

Bayes Theorem - Learning

- P(X|H) *Likelihood* probability - the evidence X conditioned on hypothesis H
  - Shows the confidence/probability of X given H
  - Given H is true (X is a peach) calculate probability that X is round and orange, i.e., x_1=round, x_2=orange.

- P(X) Prior probability of X
  - Represents the probability that sample is round and orange.

Bayes Theorem - Classification

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

Naïve Bayes Classification

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

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

- Therefore:

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

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

Naïve Bayes Classification

- 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 Bayes Classification

- Having precalculated all $P(x_k/C_i)$, an unknown example X is classified as follows:

1. For all classes calculate $P(C_i/X)$
2. Assign X to the class with the highest $P(C_i/X)$

Play Tennis?

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

Play Tennis Example: estimating $P(x_i/C)$

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 $P(C_i/x_i)$

- 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= 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 *p* (play) for example X.