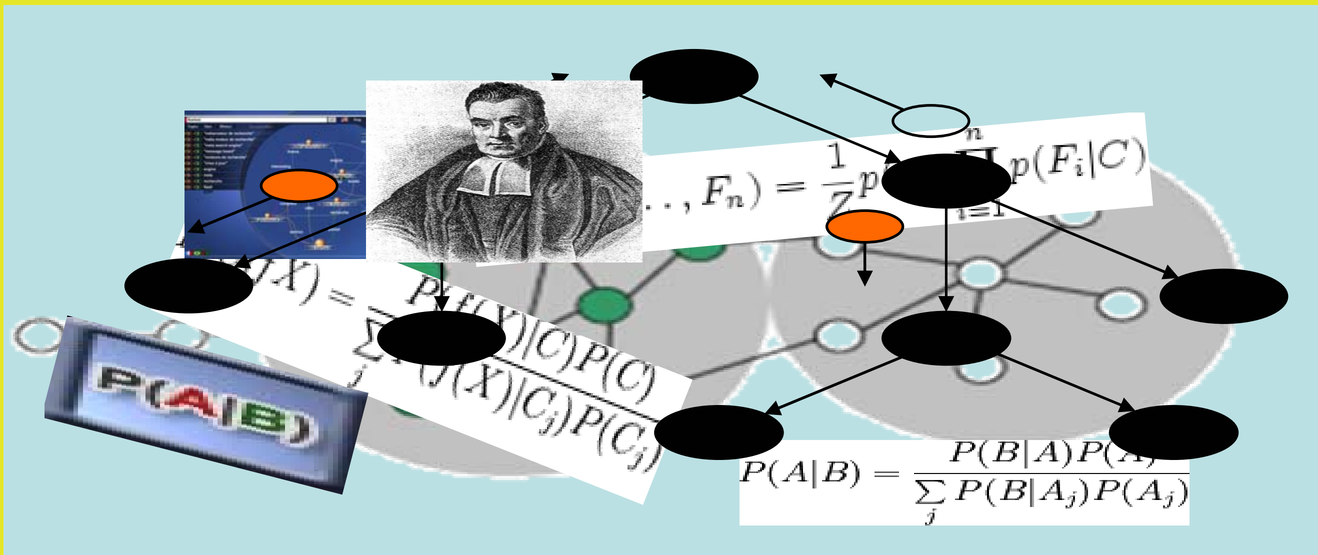
Statistical Data Mining



Application of Bayesian Statistics in Classification

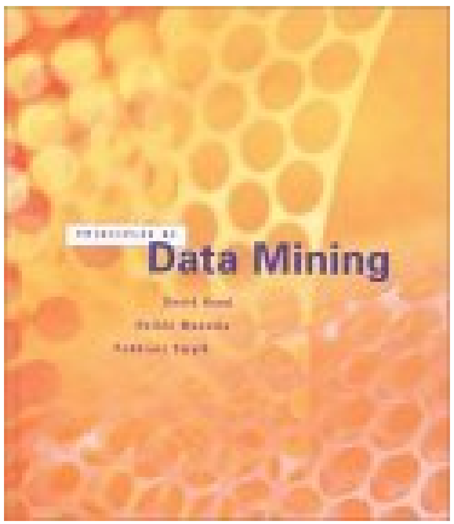
Naïve Bayes

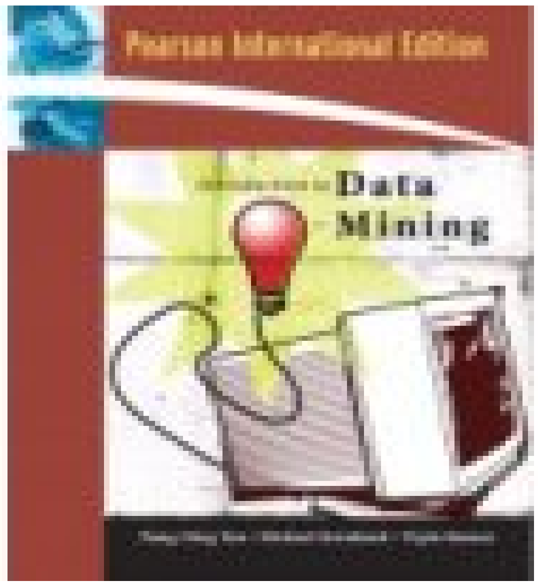
Professor Dr. Gholamreza Nakhaeizadeh

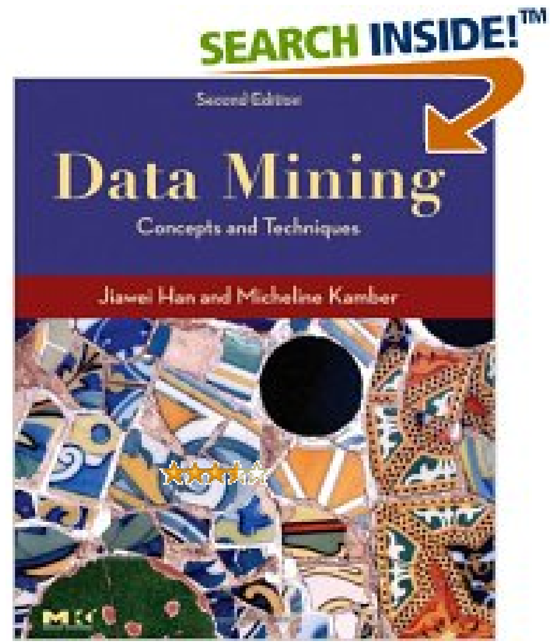
Content

- Literature used
- Introduction
- Bayes Theorem
- Application of Bayes-Theorem in classification
- Naïve Bayes

Literatur used







Principles of Data Mining
David J. Hand, Heikki Mannila, Padhraic Smyth

Pang-Ning Tan, Michael Steinbach, Vipin Kumar

Jiawei Han and Micheline Kamber

Naïve Bayes

Introduction

- In classification tasks, assigning the class value to an observation is often stochastic and not deterministic

Reasons:

- Noisy data
- Some of the relevant attributes are not considered that are stochastic

To handle such situations, one needs stochastic approaches

- Naïve Bayes
- Logistic Regression
- .....

Example:

income	car	gender	class
2000	yes	F	good
2000	yes	F	bad

Naïve Bayes

Bayes-Theorem

X and Y : Random variables

P (X,Y) : joint probability of X and Y

P(XIY)   : Conditional probability of X given Y

Then: P(X,Y) = P(YIX) . P(X) = P(XIY) . P(Y)

$$P\,(YIX) = \frac{P(XIY)\;.\;P\,(Y)}{P\,(X)} \qquad (1)$$

Relation (1) is known as Bayes-Theorem

Naïve Bayes

Application of Bayes-Theorem in classification

X : Attributes Vector,   Y: Class Vector

X and Y : Random variables

P (YIX) : Posterior probability, P(Y) : Prior probability

Bayesian Classification Task

- Building the Classifier:
  Learning P(YIX) by using data on X and Y
- Classification of new tuples:
  To each new tuple X´ assign the class value that maximizes p (Y´IX)

Naïve Bayes

Application of Bayes-Theorem in classification

How can we compute P (YIX) ?

By using the Bayes-Theorem

$$P(Y|X) = \frac{P(X|Y)\,.\,P(Y)}{P(X)}$$

- P(x) is independent of Y and can be ignored
- Computing of P(Y) can be done easily by using the observations on Y
- To compute P (XIY) there are different alternative:

Naïve Bayes is one of them

Naïve Bayes

Naïve Bayes Classifier

- Goal: Estimating the class conditional probability
- (Naïve) Assumption: given the class label Y the attributes are conditionally independent

$$P(X|Y=y) = \prod_{i=1}^{m} P(X_i \mid Y=y), \quad X = (X_1, X_2, \ldots X_m) \quad (2)$$

Instead of computing the joint conditional probability of X, it is just necessary to compute the probability of each X_i given Y

Thus, from (1) and (2) we have

$$P(Y|X) = \frac{P(Y) \cdot \prod_{i=1}^{m} P(X_i \mid Y)}{P(X)} \quad (3)$$

Naïve Bayes Classification Rule:
Assign to the new Vector X´ the class that maximizes the numerator of (3)

Naïve Bayes

Naïve Bayes Classifier

$$P(Y|X) = \frac{P(Y) \cdot \prod_{i=1}^{m} P(X_i | Y)}{P(X)}$$

$\alpha = 1/P(X) \longrightarrow$ constant

$$P(Y|X) = \alpha \; P(Y) \cdot \prod_{i=1}^{m} P(X_i | Y) \qquad (4)$$

Naïve Bayes

Example

Source: http://www-users.itlabs.umn.edu/classes/Spring-2006/csci5523/index.php?page=lecture%20slides

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Y = Evade
X= (Refund, Marital Status, Taxable Income)

New Record X':
(Refund= no,
marital status=married,
taxable income = $ 120 K)

Based on the training data, we compute

P (yes I X') and P(no I X')

The new record is classifies as "yes" if

P (yes I X') > P(no I X')

Otherwise it is classified as "no"

Naïve Bayes

Estimation conditional probabilities P (XIY)

A. Attribute is nominal

1. Choose a value of Y
2. Determine the values of the nominal attribute X that corresponds to this selected value of Y
3. Determine the fraction of these values

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Example:

1. We choose Y= Evade = No

2. The values of the attribute "Refund" that corresponds to Evade=no are:

Refund	Evade
Yes	No
No	No
No	No
Yes	No
No	No
Yes	No
No	No

3. Determining of the fractions

P (X = Yes I Y= No) = 3/7

P (X = No I Y = No) = 4/7

Naïve Bayes

Estimation conditional probabilities P (XIY)

B. Attribute is continuous-valued

Alternative 1:
Discretization of the continuous-valued Attribute. The rest of the procedure is similar to the case A

Alternative 2:
Assume a certain conditional distribution for the cotinuous-valued attribute(e.g. normal distribution)

$$P(X_i \mid Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(X_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

The distributions parameters can be estimated by using the observations on X and Y

Naïve Bayes

Estimation conditional probabilities P (XIY)

B. Attribute is continuous-valued

Example:

Taxable Income	Evade
125	No
100	No
70	No
120	No
60	No
220	No
75	No

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

$\overline{X} = (125 + 100 + 70 + \ldots + 75) / 7 = 110$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$S^2 = 17850 / 6 = 2975$$

$$S = \sqrt{2975} = 54.54$$

$$P(X_i \mid Y = \text{No}) = \frac{1}{\sqrt{2\pi}\,.\,54.54} e^{-\frac{(X_i - 110)^2}{2.\ 2975}}$$

$X_i = 120 \longrightarrow P(X_i \mid Y = \text{No}) = 0.0072$

Source: http://www-users.itlabs.umn.edu/classes/Spring-2006/csci5523/index.php?page=lecture%20slides

Naïve Bayes

Example:

Determine the class of a new Record X:
(Refund= no, marital status=married, taxable income = $ 120 K)

P (Y= No) = 7/10   P (Y = Yes) = 3/10

P (XI Y=No) =
P (Refund=No I Y=No). P(status=married I Y=No). P (T. Income = 120 IY=No)

Refund	Evade
Yes	No
No	No
No	No
Yes	No
No	No
Yes	No
No	No

→ **P (Refund = No I Y= No) = 4/7**

P (T. Income = 120 I Y= No) = 0.0072

(see the last slide)

M. status	Evade
single	No
married	No
single	No
married	No
married	No
divorced	No
married	No

→ **P (Status = married I Y= No) = 4/7**

P (X I Y=No) = 4/7 . 4/7 . 0.0072 = 0.0024

From (4) we have: P (No I X) = α   P (Y= No) . P (X I Y= No) = 7/10 . 0.0024 α= 0.0016 α

Using the same method → P (Yes I X) = 0 → P (No I X) > P (Yes I X)

The class value of the new record is computed as No

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Naïve Bayes

Strength and Weakness

- **Robust to noise and irrelevant attributes**
- **Independence assumption may not hold for some attributes**