جامعة الطائف
TAIF UNIVERSITY

DATA MINING

Assoc. Prof. Dr. Salha Alzahrani

College of Computers and Information Technology
Taif University
Saudi Arabia

s.zahrani@tu.edu.sa

Data Processing |

Recap of Previous Lecture

Content of This Lecture

Summary & Checklist

Recap of Lecture 2

- Standard Formulation
- Types of Variables
- Categorical and Continuous Attributes
- Data Preparation
  - Data Cleaning
  - Missing Values
  - Attributes Reduction

Recap of Previous Lecture

Content of This Lecture

Summary & Checklist

Content of Lecture 3

- What is Classification?
- Naïve Bayes Classifiers
- Probability of an event
- The train example
- The prior probability
- The conditional (or posterior) probability
- Naïve Bayes Algorithm
- Naïve Bayes Algorithm: The train example
- Naïve Bayes Algorithm: classification of unseen instance
- Naïve Bayes Algorithm: summary of steps
- Self-assessment Exercise
- Summary & Checklist.

Classification | What is Classification?

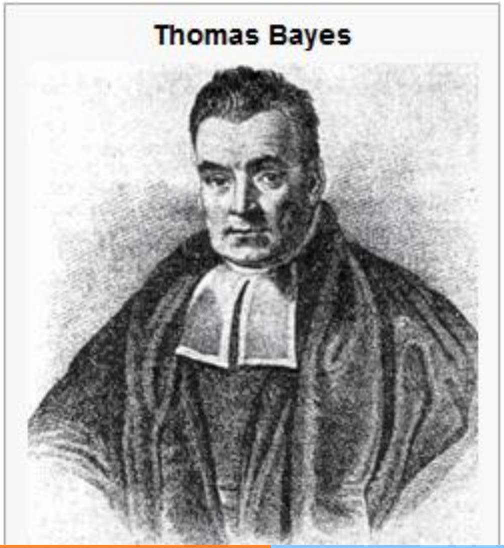
Classification is the process of dividing up objects so that each is assigned to one of a number of '***mutually exhaustive and exclusive***' categories known as **classes**.

- Classification is normally used with **labelled datasets**.
- The term ***'mutually exhaustive and exclusive'*** simply means that each object must be assigned to precisely one class, i.e. never to more than one and never to no class at all.

Classification | Naïve Bayes Classifiers

Naïve Bayes Classifiers is a method of classification that does not use rules, a decision tree or any other explicit representation of the classifier. It uses the **probability theory** to find the most likely classification of an **unseen (unclassified) instance**.

- **Naïve** = simple or straightforward
- **Bayes =** Reverend Thomas Bayes (1702–1761), an English Mathematician.



Classification | Probability of an event

The probability of an event is the number of times we would expect an **event** to occur over a long series of **trails**.

- For Example, the probability that the 6.30 p.m. train from London to Oxford arrives **on time**.
- The probability is a number from **0 to 1** inclusive, with **0** indicating ‘**impossible**’ and **1** indicating ‘**certain**’.

Classification | The train example

- In the train example, we may have four **mutually exclusive and exhaustive** events (**E**):
  - E1 – train cancelled
  - E2 – train ten minutes or more late (very late)
  - E3 – train less than ten minutes late (late)
  - E4 – train on time.

- The **probability of an event P(E)**:
  - P(E1) = 0.05
  - P(E2) = 0.1
  - P(E3) = 0.15
  - P(E4) = 0.7

- **P(E) ϵ [0,1]**
- The sum of the four probabilities has to be 1.

$$\mathbf{P(E1) + P(E2) + P(E3) + P(E4) = 1}$$

Classification | The train example

- For classification tasks, the labelled dataset is called a ***training set***.
- Each row of the training set is called an ***instance***. An instance comprises the values of a number of attributes and the corresponding classification.
- The training set constitutes the results of a sample of ***trials*** that we can use to predict the classification of other (unclassified) instance.
- Prediction Question?

weekday	winter	high	heavy	????

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

Figure 2.1 The *train* Dataset

Classification | The prior probability

The prior probability is the probability of an event based on its observed frequency in a series of trials **without any additional information**.

$$P(class = c) = \frac{frequency\ of\ c}{total\ number\ of\ instances}$$

Example:

- The prior probability of the train being *on time*:

$$P(\text{class} = \text{on time}) = 14/20 = 0.7$$

- If we have no other information, this is the best we can do.
- If we have other (relevant) information, the position is different.

Classification | The conditional (or posterior) probability

The conditional probability (also called posterior probability) is the probability of an event based on its observed frequency in a series of trials and given that **we have additional information**.

$$\boldsymbol{P(class = c \mid a = v) = \frac{frequency\ of\ c\ given\ attribute\ a}{total\ number\ of\ instances\ in\ a}}$$

Example:

P(class = on time | season = winter) is the probability that the class is *on time* given that the season is *winter*.

Classification | The conditional (or posterior) probability

The train example:

- P(class = on time | season = winter) = 2/6 = 0.33
- P(class = late | season = winter) = 2/6 = 0.33
- P(class = very late | season = winter) = 2/6 = 0.33
- P(class = cancelled | season = winter) = 0/6 = 0

Classification | Naïve Bayes Algorithm

The Naïve Bayes algorithm combines the prior probability and conditional probabilities in a single formula, to calculate the **probability of alternative classifications** .

- The method uses prior probability as before.
- But posterior probabilities are the other way round from before.

Example:

We use the conditional probability that the season is *winter* given that the class is *on time*, i.e. P(season = winter | class = on time).

Classification | Naïve Bayes Algorithm

Naïve Bayes Classification

Given a set of k mutually exclusive and exhaustive classifications $c_1, c_2, \ldots,$ c_k, which have prior probabilities $P(c_1), P(c_2), \ldots, P(c_k)$, respectively, and n attributes $a_1, a_2, \ldots, a_n$ which for a given instance have values $v_1, v_2,$ $\ldots, v_n$ respectively, the posterior probability of class c_i occurring for the specified instance can be shown to be proportional to

$$P(c_i) \times P(a_1 = v_1 \text{ and } a_2 = v_2 \ldots \text{ and } a_n = v_n \mid c_i)$$

Making the assumption that the attributes are independent, the value of this expression can be calculated using the product

$$P(c_i) \times P(a_1 = v_1 \mid c_i) \times P(a_2 = v_2 \mid c_i) \times \ldots \times P(a_n = v_n \mid c_i)$$

We calculate this product for each value of i from 1 to k and choose the classification that has the largest value.

Figure 2.3 The Naïve Bayes Classification Algorithm

Classification | Naïve Bayes Algorithm: The train example

For the *train* dataset, we can calculate all the conditional and prior probabilities, as shown in this table.

	class = on time	class = late	class = very late	class = cancelled
day = weekday	**9/14 = 0.64**	**1/2 = 0.5**	**3/3 = 1**	**0/1 = 0**
day = saturday	2/14 = 0.14	1/2 = 0.5	0/3 = 0	1/1 = 1
day = sunday	1/14 = 0.07	0/2 = 0	0/3 = 0	0/1 = 0
day = holiday	2/14 = 0.14	0/2 = 0	0/3 = 0	0/1 = 0
season = spring	4/14 = 0.29	0/2 = 0	0/3 = 0	1/1 = 1
season = summer	6/14 = 0.43	0/2 = 0	0/3 = 0	0/1 = 0
season = autumn	2/14 = 0.14	0/2 = 0	1/3 = 0.33	0/1 = 0
season = winter	**2/14 = 0.14**	**2/2 = 1**	**2/3 = 0.67**	**0/1 = 0**
wind = none	5/14 = 0.36	0/2 = 0	0/3 = 0	0/1 = 0
wind = high	**4/14 = 0.29**	**1/2 = 0.5**	**1/3 = 0.33**	**1/1 = 1**
wind = normal	5/14 = 0.36	1/2 = 0.5	2/3 = 0.67	0/1 = 0
rain = none	5/14 = 0.36	1/2 =0.5	1/3 = 0.33	0/1 = 0
rain = slight	8/14 = 0.57	0/2 = 0	0/3 = 0	0/1 = 0
rain = heavy	**1/14 = 0.07**	**1/2 = 0.5**	**2/3 = 0.67**	**1/1 = 1**
Prior Probability	**14/20 = 0.70**	**2/20 = 0.10**	**3/20 = 0.15**	**1/20 = 0.05**

Figure 2.2 Conditional and Prior Probabilities: *train* Dataset

Classification | Naïve Bayes Algorithm: classification of unseen instance

Using the values in each of the columns in the table, we can obtain the posterior probabilities for **each possible classification** for the unseen instance:

weekday	winter	high	heavy	????

- **class = on time**

  $0.70 \times 0.64 \times 0.14 \times 0.29 \times 0.07 = 0.0013$
- **class = late**

  $0.10 \times 0.50 \times 1.00 \times 0.50 \times 0.50 = 0.0125$
- **class = very late**

  $0.15 \times 1.00 \times 0.67 \times 0.33 \times 0.67 = 0.0222$
- **class = cancelled**

  $0.05 \times 0.00 \times 0.00 \times 1.00 \times 1.00 = 0.0000$

- **The largest value is for class = very late**

Naïve Bayes Algorithm

Step 1: Define the classes, and possible values for each attribute.

Step 2: Calculate the prior probability of each class.

Step 3: Calculate the posterior probability of each attribute given that each class (in a table).

Step 4: Using the values in each of the columns in the table, calculate the posterior (conditional) probabilities for **each possible classification** for a given unseen (unclassified) instance.

Step 5: Choose the largest value as a classification of the given unseen instance.

Classification | Self-assessment Exercise

- Using the Naïve Bayes classification algorithm with the *train* dataset, calculate the most likely classification for the following unseen instances.

weekday	summer	high	heavy	????

sunday	summer	normal	slight	????

Recap of Previous Lecture

Content of This Lecture

Summary & Checklist

Summary & Checklist

- ☑ What is Classification?
- ☑ Naïve Bayes Classifiers
- ☑ Probability of an event
- ☑ The train example
- ☑ The prior probability
- ☑ The conditional (or posterior) probability
- ☑ Naïve Bayes Algorithm
- ☑ Naïve Bayes Algorithm: The train example
- ☑ Naïve Bayes Algorithm: classification of unseen instance
- ☑ Naïve Bayes Algorithm: summary of steps
- ☑ Self-assessment Exercise

Reminder | Next Lecture !

Next Lecture…

Classification using Nearest Neighbour Algorithm (Ch. 2)

- *Be ready!*
- *Do your self-assessment exercise.*
- *Prepare your course portfolio.*
- *Download & print the lecture notes before your class.*

Thank You !

Q&A

✉ s.zahrani@tu.edu.sa