



Machine Learning: Classification Techniques

Prajwol Sangat



程序代写代做 CS 编程辅导



MONASH
INFORMATION
TECHNOLOGY

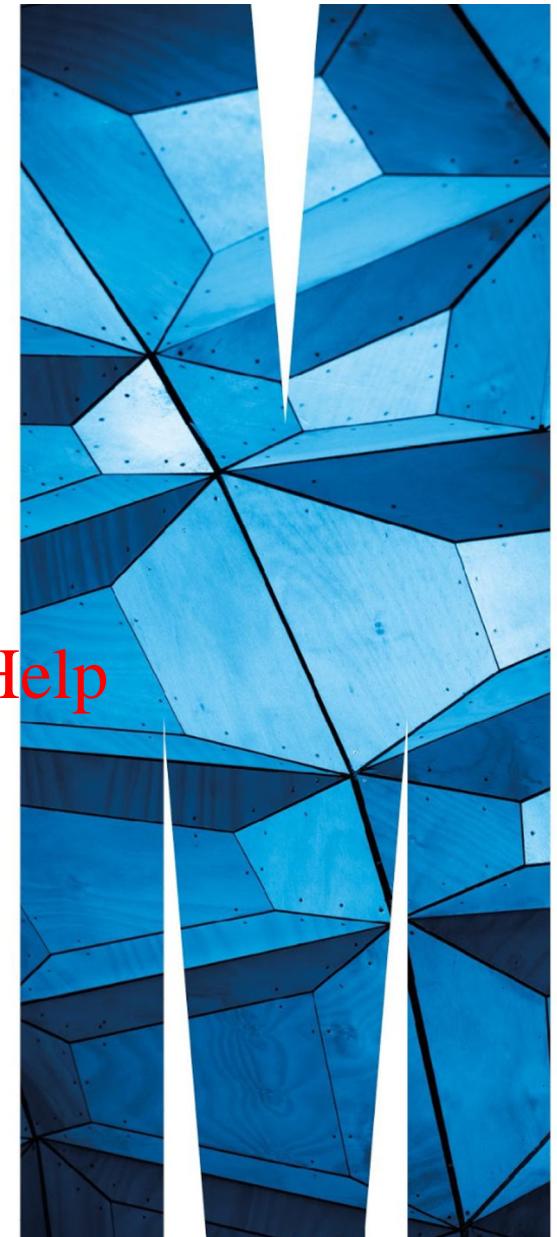
WeChat: esutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



Last week

程序代写代做 CS编程辅导

Data Transformer, E
Feature Selection and Pipelines



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

This week

程序代写代做 CS编程辅导

Classification Algorit

Decision Tree

Random Forest

DEMO



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Classification

程序代写代做 CS编程辅导

Predictive Data Model

A classifier model needs to be created using training dataset

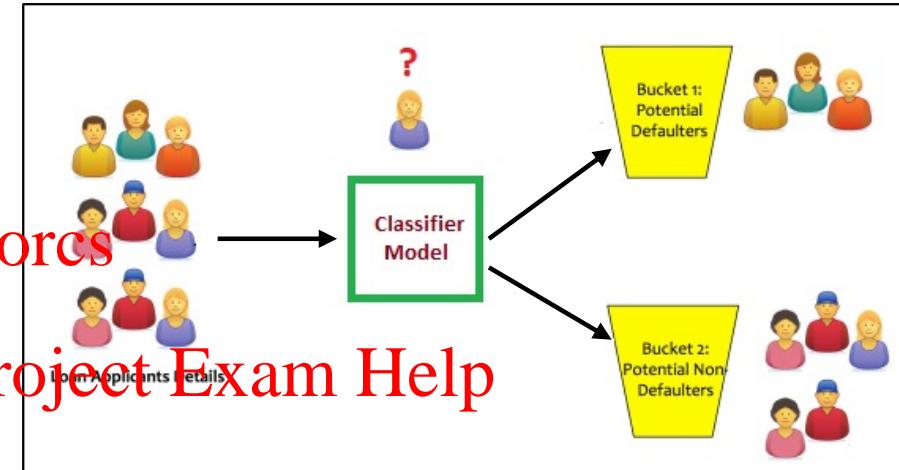
After the classifier is created,
classification is the process of
assigning new instances from
the testing dataset to
predefined classes

WeChat: cstutorcs
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



Classification

程序代写代做 CS编程辅导

Classifiers can be:

- *Binary classifier*
- *Multi-Class classifier*

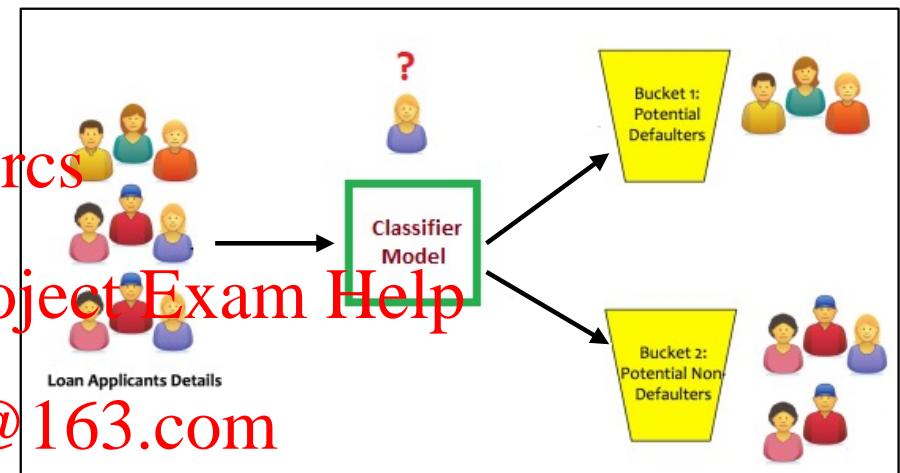


Binary classifiers: Classification with only 2 distinct classes or with 2 possible outcomes

Example: classification of spam email and non-spam email, potential defaulter and non-defaulter

Multi-Class classifiers: Classification with more than two distinct classes

Example: classification of types of animals, classification of books into categories.



<https://tutorcs.com>

Classification Algorithms

There are several types of classification algorithms in Machine Learning.

- Decision Trees
- Random Forest
- Logistic Regression
- And Many More.



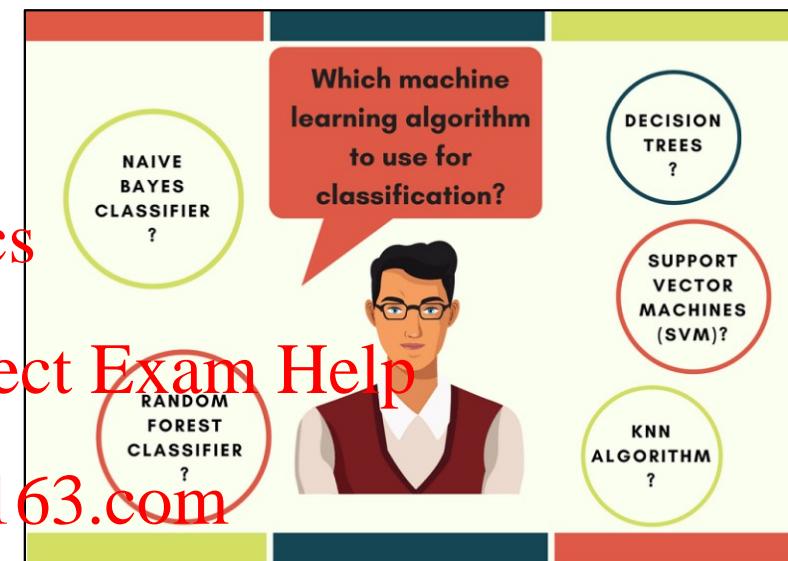
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



Decision Trees

程序代写代做 CS编程辅导

- A tree-like predictive model for classification and regression
- In DTs, a record/sample which belongs to a certain class or category is identifiable through features/attributes.
- It splits samples into two or more homogeneous sets (leaves) based on the most significant attributes (predictors)



WeChat: estutories

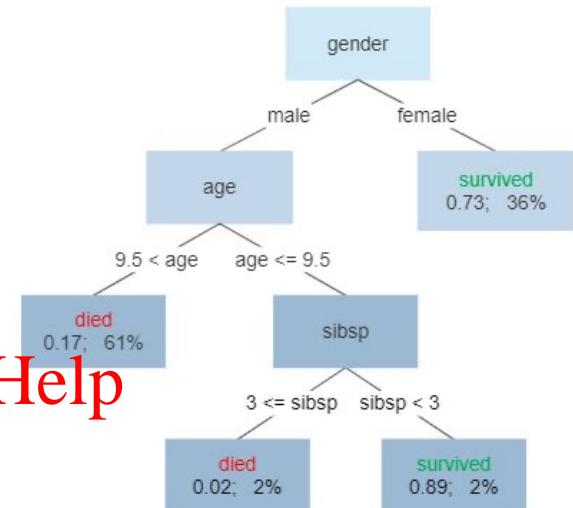
Assignment Project Exam Help

Samples	Features/Attributes			Class
	gender	age	sibsp	
Person 1	male	30	1	died
Person 2	female	20	2	survived

QQ: 749389476

<https://tutorcs.com>

Survival of passengers on the Titanic



Example: Titanic dataset

Decision Trees

程序代写代做 CS编程辅导

- Each **internal node** represents a "test" on an attribute (e.g. gender)
- Each **branch** corresponds to attribute values (outcome of test) – e.g. male or female
- Each **leaf/terminal node** assigns class label (e.g., died or survived)



WeChat: cstutorcs

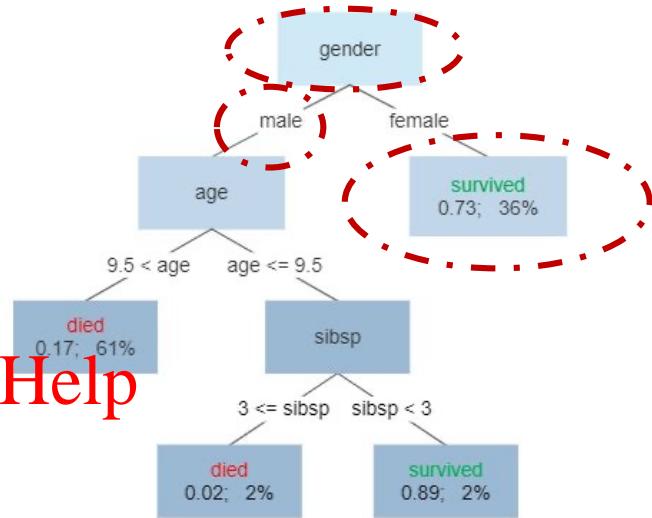
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Survival of passengers on the Titanic



Example: *Titanic dataset*

Decision Tree Algorithm



Common terms used with Decision Trees:

Root Node: It represents entire population of people and this further gets divided into two or more homogeneous groups.

Splitting: It is a process of dividing a node into two or more sub-nodes.

Decision Node: When a sub-node splits into further sub-nodes, then it is called decision node.

Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.

Pruning: When sub-nodes of a decision node is removed, this process is called pruning (an opposite process of splitting).

Branch / Sub-Tree: A sub section of entire tree is called branch or sub-tree.

Parent and Child Node: A node, which is divided into sub-nodes is called parent node of sub-nodes whereas sub-nodes are the child of parent node.

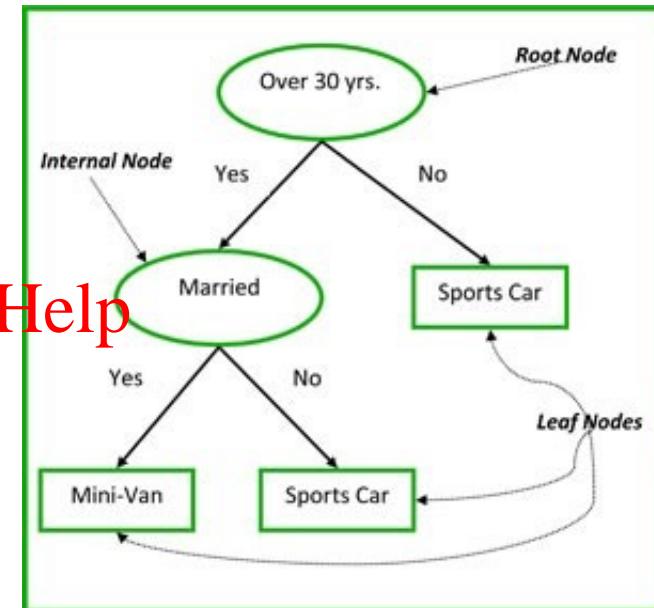
<https://tutorcs.com>

QQ: 749389476

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com



Decision Tree Algorithm



Supervised Learning – need input labels to build a DT

Constructing a DT is generally a recursive process

- Initialization: All training data at the root node
- Partition training data recursively by choosing one attribute at a time
- Repeat process for partitioned dataset
- Stopping criteria: When all training data in each partition have same target class

The most common approach in building a decision tree:

ID3 (Iterative Dichotomiser 3) → uses **Entropy function** and **Information gain** as metrics to construct a DT.

<https://tutorcs.com>

ID3 (Iterative Dichotomiser 3)



ID3 (Iterative Dichotomiser 3) was developed in 1986 by Ross Quinlan.

The algorithm creates a **multiway tree**, finding for each node (i.e. in a greedy manner) the **categorical feature that will yield the largest information gain**.

Trees are grown to their maximum size and then a pruning step is usually applied to improve the ability of the tree to generalise to unseen data.

<https://tutorcs.com>

Entropy

程序代写代做 CS编程辅导

- Measure of uncertainty in data
- Informs the predictability of event
 - Low value -> Less uncertainty, high value -> high uncertainty



$$H(S) = \sum_{i=1}^n p_i \log \frac{1}{p_i}$$

p_i - Probability of event i
n - Number of events

Less homogeneous

Play Basketball	
Yes	No
9	5

More homogeneous

Play Basketball	
Yes	No
13	1

Assignment Project Exam Help

$$= -\left(\frac{9}{14} \log \frac{9}{14}\right) - \left(\frac{5}{14} \log \frac{5}{14}\right)$$

Email: tutorcs@163.com

QQ: 749389476

$$H(Play_basketball) = -\left(\frac{13}{14} \log \frac{13}{14}\right) - \left(\frac{1}{14} \log \frac{1}{14}\right)$$
$$= 0.1115$$

<https://tutorcs.com>

If samples are completely homogeneous, the entropy is zero

Information gain

程序代写代做 CS编程辅导

- IG for a set S is change in entropy after deciding on attribute A.
- It computes difference between entropy before split and average entropy after split of the dataset based on an attribute A
- Used to decide which attributes are more relevant in ID3 algorithm



$$\begin{aligned} S, A) &= H(S) - H(S, A) \\ &= H(S) - \sum_{i \in \text{values}(A)} p_i H(S_i) \end{aligned}$$

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

S_i Subset/partition of data after splitting S

Weighted sum entropy given A

Example

程序代写代做 CS编程辅导

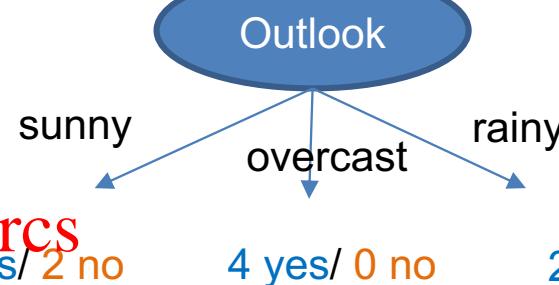
$$H(E) = p(yes) \log \frac{1}{p(yes)} + p(no) \log \frac{1}{p(no)}$$

$$= 0.2831$$

		Play Basketball		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14



9 yes/ 5 no



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

$$H(S_{sunny}) = -\left(\frac{3}{5} \log \frac{3}{5}\right) - \left(\frac{2}{5} \log \frac{2}{5}\right) = 0.2922$$

$$H(S_{overcast}) = 0$$

$$H(S_{rain}) = -\left(\frac{2}{5} \log \frac{2}{5}\right) - \left(\frac{3}{5} \log \frac{3}{5}\right) = 0.2922$$

IG(Play_basketball, outlook)

$$\begin{aligned} H(S_A) &= p(sunny)H(S_{sunny}) + p(overcast)H(S_{overcast}) + p(rain)H(S_{rain}) \\ &= \frac{5}{14}(0.2922) + \frac{4}{14}(0) + \frac{5}{14}(0.2922) \\ &= 0.2087 \end{aligned}$$

<https://tutorcs.com>

ID3 (Iterative Dichotomiser 3)

- It constructs DT, by finding node attribute that returns the highest information gain to split

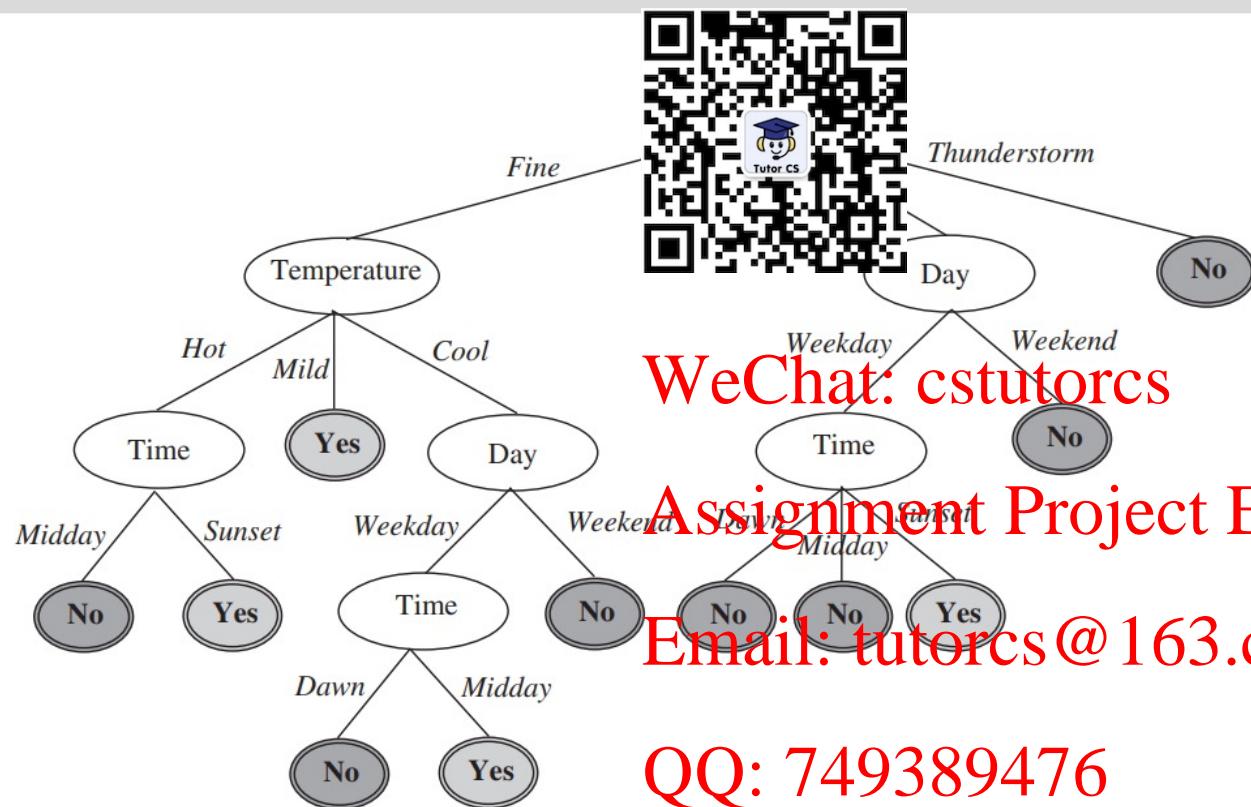


Steps

- Compute the entropy for dataset S $\rightarrow H(S)$
- For every attribute/feature A :
 - Calculate entropy for each categorical value of A $\rightarrow H(S_i)$
 - Take weighted average entropy for the current attribute $H(S, A) = \sum_{i \in Values(A)} p_i H(S_i)$
 - Calculate IG for the current attribute $\rightarrow IG(S, A) = H(S) - H(S, A)$
- Pick the attribute with highest IG to be a node, and split dataset by its branch to child nodes/subsets $\rightarrow QQ: 749389476$
- Repeat same process at every child node until the tree is complete $\rightarrow https://tutorcs.com$

Stopping condition: when data in each partition have same target class

Decision Trees: To Jog or Not To Jog



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

A decision tree is constructed based only on the given training dataset. It is not based on a universal belief.

Figure 17.10 A decision tree

ID3

程序代写代做 CS编程辅导

Example: Consider a piece of data collected over the course of 15 days where the features are Weather, Temperature, Time, Day and the outcome variable is whether Jogging was done on the day. Now, our job is to build a predictive model which takes in above 4 parameters and predicts whether Jogging will be done on the day. We'll build a decision tree to do that using **ID3 algorithm.**



	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Rainy	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

WeChat: cstutorcs
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

Figure 17.11. Training dataset

<https://tutorcs.com>

ID3 (Iterative Dichotomiser 3)

- It constructs DT, by finding node attribute that returns the highest information gain to split



Steps

- Compute the entropy for dataset S $\rightarrow H(S)$
- For every attribute/feature A :
 - Calculate entropy for each categorical value of A $\rightarrow H(S_i)$
 - Take weighted average entropy for the current attribute $H(S, A) = \sum_{i \in Values(A)} p_i H(S_i)$
 - Calculate IG for the current attribute $\rightarrow IG(S, A) = H(S) - H(S, A)$
- Pick the attribute with highest IG to be a node, and split dataset by its branch to child nodes/subsets **QQ: 749389476**
- Repeat same process at every child node until the tree is complete

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
<https://tutorcs.com>

Stopping condition: when data in each partition have same target class

ID3

程序代写代做 CS编程辅导

Entropy for the given probability of the target class, p_1, p_2, \dots, p_n where

$\sum_{i=1}^n p_i = 1$, can be calculated as follows:

$$\text{entropy}(p_1, p_2, \dots, p_n) = \sum_{i=1}^n (p_i \log(1/p_i))$$



WeChat: cstutorcs

$$\begin{aligned}\text{entropy(Yes, No)} &= 5/15 \cdot \log(15/5) + 10/15 \cdot \log(15/10) \\ &= 0.2764\end{aligned}$$

Email: tutorcs@163.com

- Step 1: Calculate entropy for the training dataset in Figure 17.11. The result is previously calculated as 0.2764 (see equation 17.3).

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Jog	
Yes	No
5	10

ID3 (Iterative Dichotomiser 3)

- It constructs DT, by finding node attribute that returns the highest information gain to split



Steps

- Compute the entropy for dataset S $\rightarrow H(S)$
- For every attribute/feature A:
 - Calculate entropy for each categorical value of A $\rightarrow H(S_i)$
 - Take weighted average entropy for the current attribute $H(S, A) = \sum_{i \in Values(A)} p_i H(S_i)$
 - Calculate IG for the current attribute $\rightarrow IG(S, A) = H(S) - H(S, A)$
- Pick the attribute with highest IG to be a node, and split dataset by its branch to child nodes/subsets **QQ: 749389476**
- Repeat same process at every child node until the tree is complete

WeChat: estutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
<https://tutorcs.com>

Stopping condition: when data in each partition have same target class

ID3

程序代写代做 CS编程辅导

$$\begin{aligned} \text{entropy}(\text{Weather}=\text{Fine}) &= 4/7 \times \log(7/4) \\ &= 0.2966 \end{aligned}$$



(17.4)

$$\begin{aligned} \text{entropy}(\text{Weather}=\text{Shower}) &= 1/4 \times \log(4/1) \\ &= 0.2442 \end{aligned}$$

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Weather	Jog	
	Yes	No
Fine	4	3
Shower	1	3
Thunderstorm	0	4
		15

ID3

程序代写代做 CS编程辅导

$$\begin{aligned}\text{Weighted sum entropy (Weather)} &= \text{Weighted entropy}(Fine) + \text{Weighted entropy}(Shower) + \text{Weighted entropy}(Thunderstorm) \\ &= 7/15 \times 0.2966 + 4/15 \times 0.2442 + 4/15 \times 0.2442 \\ &= 0.2035\end{aligned}$$



(17.6)

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Weather	Jog		15
	Yes	No	
Fine	4	3	7
Shower	1	3	4
Thunderstorm	0	4	4

ID3

程序代写代做 CS编程辅导

$$\begin{aligned} \text{gain}(\text{Weather}) &= \text{entropy}(\text{training dataset } D) \\ &= 0.2764 - 0.2035 \\ &= 0.0729 \end{aligned}$$



attribute Weather)

(17.7)

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

- Step 2: Process attribute *Weather*

- Calculate weighted sum entropy of attribute *Weather*:

$$\text{entropy}(\text{Fine}) = 0.2966 \quad (\text{equation 17.4})$$

$$\text{entropy}(\text{Shower}) = 0.2442 \quad (\text{equation 17.5})$$

$$\text{entropy}(\text{Thunderstorm}) = 0 + 4/4 \times \log(4/4) = 0$$

$$\text{weighted sum entropy}(\text{Weather}) = 0.2035 \quad (\text{equation 17.6})$$

- Calculate information gain for attribute *Weather*:

$$\text{gain}(\text{Weather}) = 0.0729 \quad (\text{equation 17.7})$$



- Step 3: Process attribute *Temperature*

- Calculate weighted sum entropy of attribute *Temperature*:

$$\text{entropy}(\text{Hot}) = 2/5 \times \log(5/2) + 3/5 \times \log(5/3) = 0.2923$$

$$\text{entropy}(\text{Mild}) = \text{entropy}(\text{Hot})$$

$$\text{entropy}(\text{Cool}) = 1/5 \times \log(5/1) + 4/5 \times \log(5/4) = 0.2173$$

$$\begin{aligned} \text{weighted sum entropy}(\text{Temperature}) &= 5/15 \times 0.2923 + 5/15 \times 0.2173 \\ &= 0.2674 \end{aligned}$$

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 7.11. Training dataset

	Temperature	Jog		15
		Yes	No	
	Hot	2	3	5
	Mild	3	2	5
	Cool	1	4	5

ID3

程序代写代做 CS编程辅导



- Step 4: Process attribute *Time*

- Calculate weighted sum entropy of attribute *Time*:

$$\text{entropy}(\text{Dawn}) = 0 + 5/5 \times \log(5/5) = 0$$

$$\text{entropy}(\text{Midday}) = 2/6 \times \log(6/2) + 4/6 \times \log(6/4) = 0.2764$$

$$\text{entropy}(\text{Sunset}) = 3/4 \times \log(4/3) + 1/4 \times \log(4/1) = 0.2443$$

$$\text{weighted sum entropy } (\text{Time}) = 0 + 6/15 \times 0.2764 + 4/15 \times 0.2443 = 0.1757$$

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

25

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 7.11. Training dataset

		Jog	
		Yes	No
Time	Dawn	0	5
	Midday	2	4
	Sunset	3	1
			15

ID3

程序代写代做 CS编程辅导



Step 5: Process attribute *Day*

- Calculate weighted sum entropy of attribute *Day*:

$$\begin{aligned} \text{entropy(Weekday)} &= 4/10 \times \log(10/4) + 6/10 \times \log(10/6) \\ &= 0.2923 \end{aligned}$$

$$\begin{aligned} \text{entropy(Weekend)} &= 1/5 \times \log(5/1) + 4/5 \times \log(5/4) \\ &= 0.2173 \end{aligned}$$

$$\begin{aligned} \text{weighted sum entropy (Day)} &= 10/15 \times 0.2923 + 5/15 \\ &\quad \times 0.2173 = 0.2674 \end{aligned}$$

- Calculate information gain for attribute *Day*:

$$\text{gain}(\text{Temperature}) = 0.2764 - 0.2674 = 0.009$$

QQ: 749389476

<https://tutorcs.com>

WeChat: cstutorcs

Email: tutorcs@163.com

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 7.11. Training dataset

Day	Jog	
	Yes	No
Weekend	4	6
Weekday	1	4
		15

ID3 (Iterative Dichotomiser 3)

- It constructs DT, by finding node attribute that returns the highest information gain to split



Steps

- Compute the entropy for dataset S

$$\rightarrow H(S)$$

- For every attribute/feature A :

- Calculate entropy for each categorical value of A

$$H(S_i)$$

- Take weighted average entropy for the current attribute

$$H(S, A) = \sum_{i \in Values(A)} p_i H(S_i)$$

- Calculate IG for the current attribute

$$IG(S, A) = H(S) - H(S, A)$$

WeChat: cstutorcs

Assignment Project Exam Help

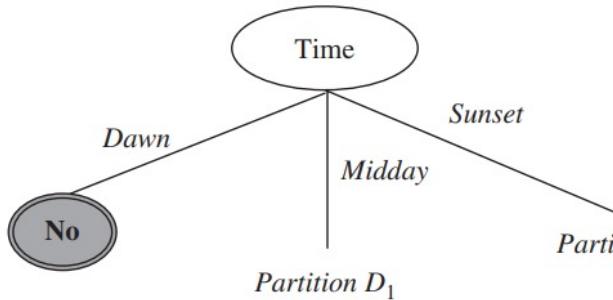
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

ID3

程序代写代做 CS编程辅导



WeChat: cstutorcs
Figure 17.13. Attribute *Time* as the root node

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Comparing equations 17.7, 17.8, 17.9, and 17.10 ,and 17.10 for the gain of each other attributes (Weather, Temperature, Time, and Day), the biggest gain is *Time*, with gain value = 0.1007 (see equation 17.9), and as a result, attribute *Time* is chosen as the first splitting attribute. A partial decision tree with the root node *Time* is shown in Figure 17.13.

<https://tutorcs.com>

ID3

程序代写代做 CS编程辅导



		Jog		
		Yes	No	
Day	Weekend	0	0	0
	Weekday	2	4	6
				6

WeChat: cstutorcs

- The next stage is to process partition D_1 consisting of records with ~~Time=Midday~~. Training dataset partition D_1 consists of 6 records with record#: 3, 6, 8, 9, 10, and 15. The next task is to determine the splitting attribute for partition D_1 , whether it is ~~Weather, Temperature, or Day~~.

<https://tutorcs.com>

		Yes	No	
Weather	Fine	2	1	3
	Shower	0	1	1
	Thunderstorm	0	2	2
				6

Temperature	Yes	No		
	Hot	0	2	2
	Mild	1	1	2
Cool	1	1	2	
				6

程序代写代做 CS编程辅导

Step 1: Calculate entropy for the training dataset partition D_1 .

$$\text{entropy}(D_1) = 2/6 \log(6/2) + 4/6 \log(6/4)$$



Step 2: Process attribute *Weather*

- Calculate weighted sum entropy of attribute *Weather*:

$$\text{entropy}(\text{Fine}) = 2/3 \times \log(6/2) + 1/3 \times$$

$$\text{entropy}(\text{Shower}) = \text{entropy}(\text{Thunderstorm})$$

$$\text{weighted sum entropy} (\text{Weather}) = 3/5 \times 0.2764 = 0.1382$$

- Calculate information gain for attribute *Weather*:

$$\text{gain}(\text{Weather}) = 0.2764 - 0.1382 = 0.1382 \quad (17.12)$$

Step 3: Process attribute *Temperature*

- Calculate weighted sum entropy of attribute *Temperature*:

$$\text{entropy}(\text{Hot}) = 0$$

$$\text{entropy}(\text{Mild}) = \text{entropy}(\text{Cool}) = 1/2 \times \log(2/1) + 1/2$$

$$\times \log(2/1) = 0.3010$$

$$\text{weighted sum entropy} (\text{Temperature}) = 2/6 \times 0.3010 + 2/6$$

$$\times 0.3010 = 0.2006$$

- Calculate information gain for attribute *Temperature*:

$$\text{gain}(\text{Temperature}) = 0.2764 - 0.2006 = 0.0758 \quad (17.13)$$

Step 4: Process attribute *Day*

- Calculate weighted sum entropy of attribute *Day*:

$$\text{entropy}(\text{Weekday}) = 2/6 \times \log(6/2) + 4/6 \times \log(6/4) = 0.2764$$

$$\text{entropy}(\text{Weekend}) = 0$$

$$\text{weighted sum entropy} (\text{Day}) = 0.2764$$

- Calculate information gain for attribute *Day*:

$$\text{gain}(\text{Temperature}) = 0.2764 - 0.2764 = 0 \quad (17.14)$$

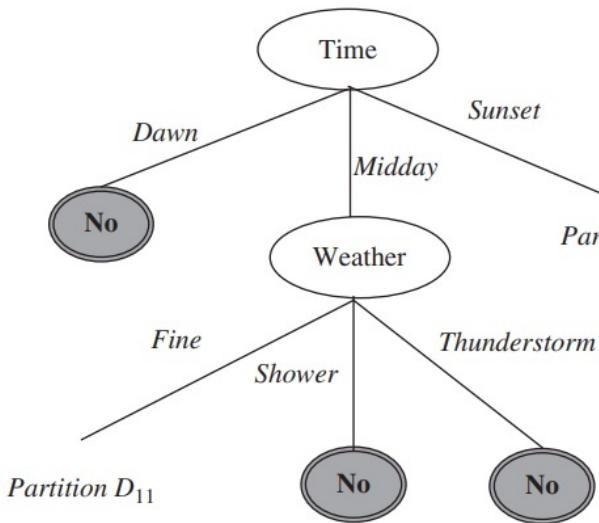
WeChat: cstutorcs

The best splitting node for partition D_1 is attribute *Weather* with information gain value of 0.1382 (see equation 17.12)

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



WeChat: cstutorcs

Assignment Project Exam Help

Figure 17.14 Attribute
Weather as next splitting attribute

- The next stage is to process partition D_1 consisting of records with Time=Midday. Training dataset partition D_1 consists of 6 records with record#s: 3, 6, 8, 9, 10, and 15. The next task is to determine the splitting attribute for partition D_1 , whether it is Weather, Temperature, or Day.

<https://tutorcs.com>

Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

ID3

程序代写代做 CS编程辅导

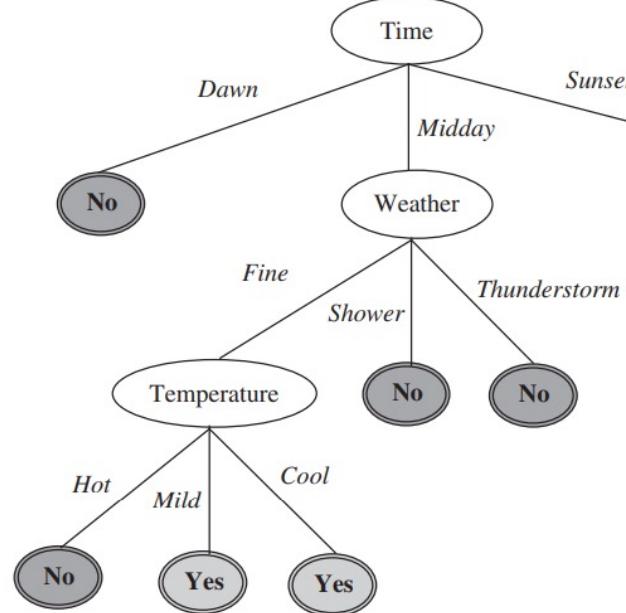


Figure 17.15 Final decision tree

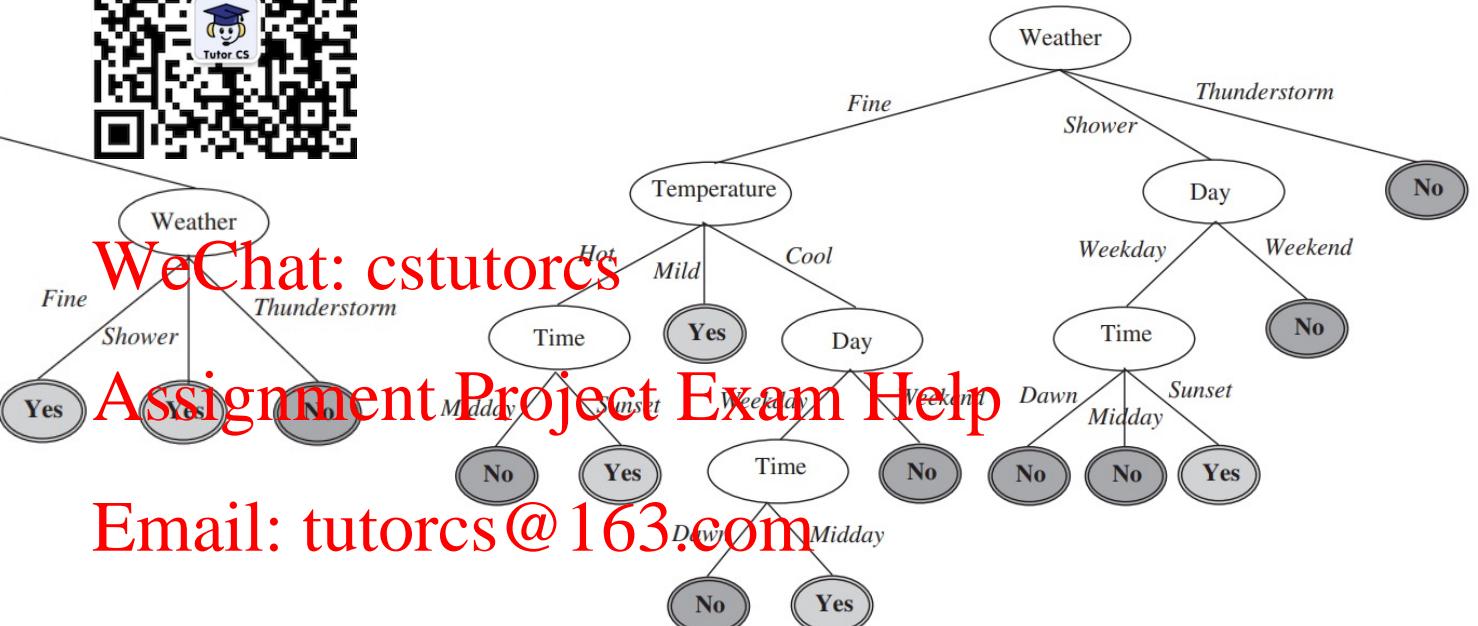


Figure 17.10 A decision tree

<https://tutorcs.com>

Maximum Depth of DT

maxDepth: the largest possible number of edges between the root to a leaf (or maximum level of the tree).



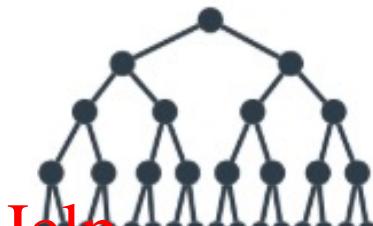
Depth = 1



Depth = 2



Depth = 3



Depth = 4

WeChat: cstutorcs
Assignment Project Exam Help

Email: tutorcs@163.com

Maximum depth of a decision tree
QQ: 749389476

<https://tutorcs.com>

Pruning

程序代写代做 CS编程辅导



<https://kaumadiechamalka100.medium.com/decision-tree-in-machine-learning-c610ef087260>

<https://tutorcs.com>

Decision Tree Algorithm



Advantages:

- Easy to understand.
- Easy to generate rules.
- There are almost null hyper-parameters to be tuned.
- Complex Decision Tree models can be significantly simplified by its visualizations.

WeChat: cstutorcs

Disadvantages:

Assignment Project Exam Help

- Might suffer from overfitting.
- Does not easily work with non-numerical data.
- Low prediction accuracy for a dataset in comparison with other machine learning classification algorithms.
- When there are many class labels, calculations can be complex.

Email: tutorcs@163.com
QQ: 749389476

<https://tutorcs.com>

Ensemble methods



A single decision tree has the tendency to overfit

But, it is super fast

How about multiple trees at once?

WeChat: cstutorcs
Assignment Project Exam Help

Make sure they do not all just learn the same!
Email: tutorcs@163.com

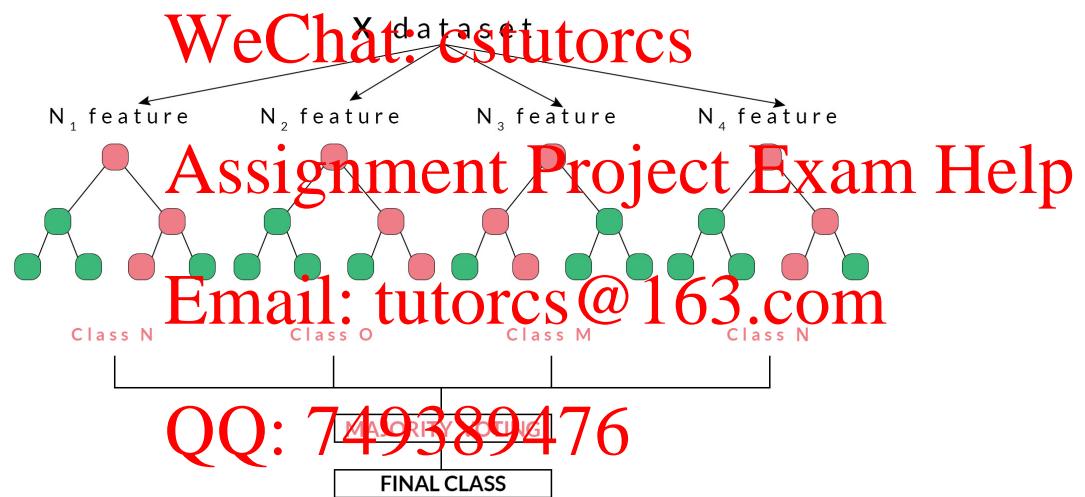
QQ: 749389476

<https://tutorcs.com>

Random Forest Algorithm



Random forest (or **forests**) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's outputs by individual trees.



<https://tutorcs.com>

Optimisations

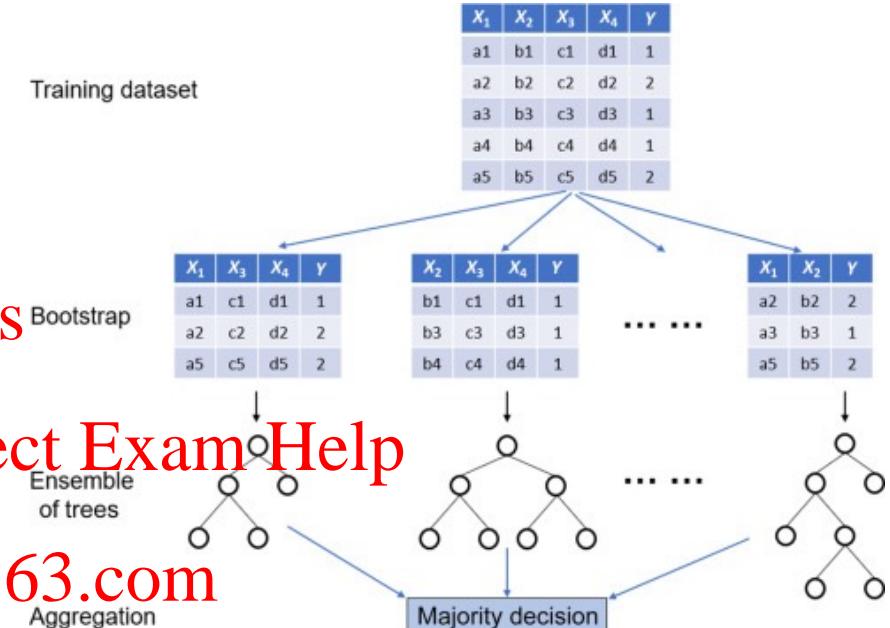
程序代写代做 CS编程辅导

1. Bagging: Bootstrap aggregating is a method that result in low variance - used to reduce variance of D.T.s



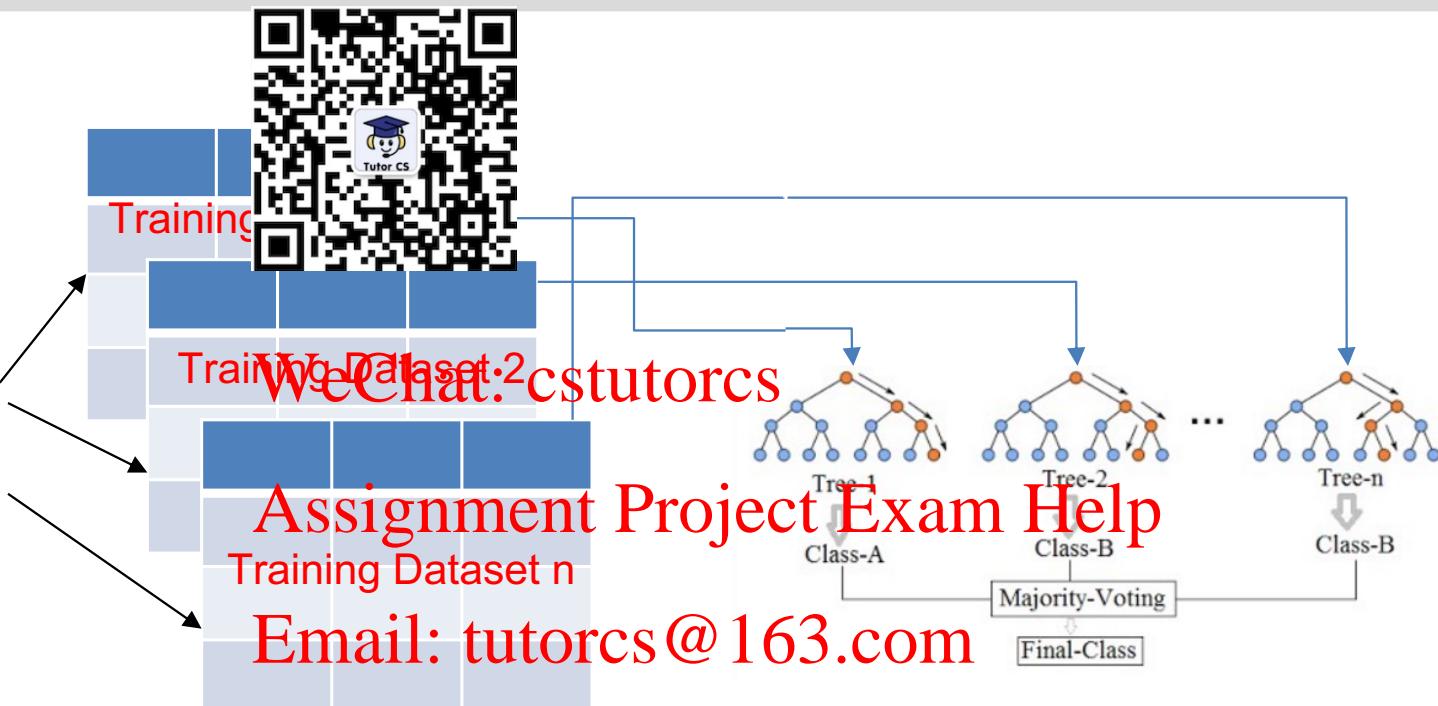
Rather than training each tree on all the inputs in the training set (producing multiple identical trees), each tree is trained on different set of sample data.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476



Example

程序代写代做 CS编程辅导



Optimisations

程序代写代做 CS编程辅导

2. Gradient boosting



Combining best classifiers to improve prediction accuracy by adding each new tree.

- It works by combining several weak learners (typically high bias, low variance models) to produce an overall strong model.
- It builds one tree at a time, works in a forward stage-wise manner, - adding a classifier at a time, so that the next classifier is trained to improve the already trained ensemble.

QQ: 749389476

<https://tutorcs.com>

Advantages and Disadvantages of Random Forest



Advantages

- It is robust to correlated predictors.
- It is used to solve both regression and classification problems.
- It can be also used to solve unsupervised ML problems.
- It can handle thousands of input variables without variable selection.
- It can be used as a feature selection tool using its variable importance plot.
- It takes care of missing data internally in an effective manner.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Advantages and Disadvantages of Random Forest



Disadvantages

- The Random Forest model is difficult to interpret.
- It tends to return erratic predictions for observations out of range of training data. For example, the training data contains two variable x and y. The range of x variable is 30 to 70. If the test data has x = 200, random forest would give an unreliable prediction.
- It can take longer than expected time to compute a large number of trees.

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

What have we learnt today?



Classification techniques

Decision Trees and Random Forest and KNN

When and how to use each technique

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



MONASH
University

程序代写代做 CS 编程辅导

MONASH INFORMATION TECHNOLOGY



Parallel Classification

Assignment Project Exam Help

Email: tutorcs@163.com

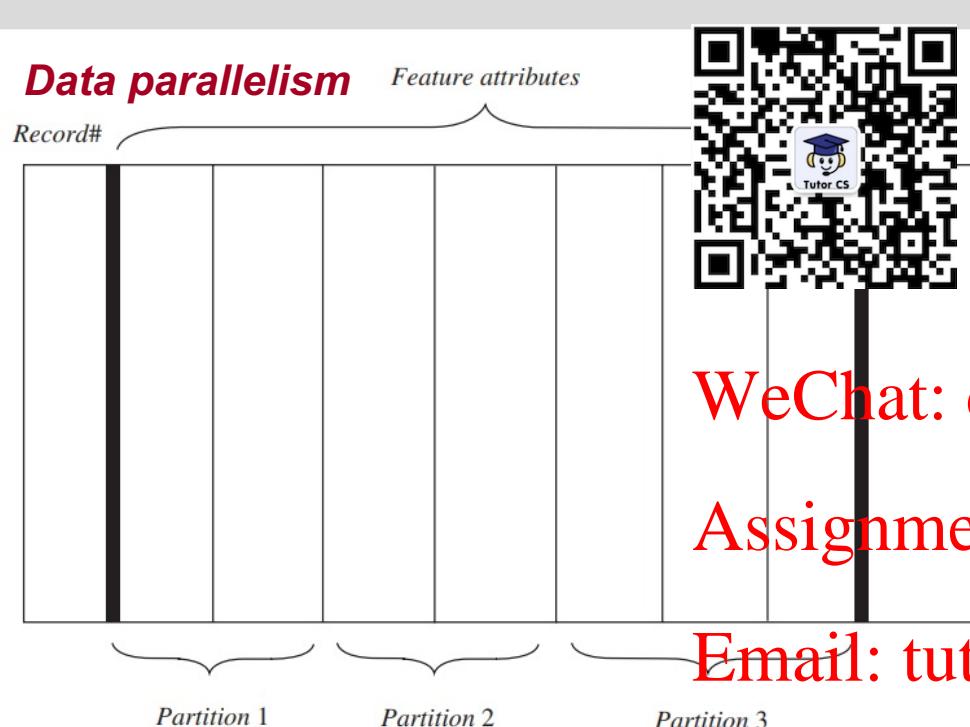
QQ: 749389476

<https://tutorcs.com>

Prajwol Sangat



Parallel Classification: Decision Tree



Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Figure 17.16 Vertical data partitioning of training data set

QQ: 749389476

<https://tutorcs.com>

Parallel Classification: Decision Tree

Data parallelism: Vertical Partitioning dataset

Rec#	Weather	Temperature	Class
1	Fine	Mild	No
2	Fine	Hot	No
3	Shower	Mild	No
4	Thunderstorm	Cool	No
5	Shower	Hot	Yes
6	Fine	Hot	No
7	Fine	Cool	No
8	Thunderstorm	Cool	No
9	Fine	Cool	Yes
10	Fine	Mild	Yes
11	Shower	Hot	No
12	Shower	Mild	No
13	Fine	Cool	No
14	Thunderstorm	Mild	No
15	Thunderstorm	Hot	No



Rec#	Time	Day	Jog (Target Class)
1	Sunset	Weekend	Yes
2	Sunset	Weekday	Yes
3	Midday	Weekday	No
4	Dawn	Weekend	No
5	Sunset	Weekday	Yes
6	Midday	Weekday	No
7	Dawn	Weekend	No
8	Midday	Weekday	No
9	Midday	Weekday	Yes
10	Midday	Weekday	Yes
11	Dawn	Weekend	No
12	Dawn	Weekday	No
13	Dawn	Weekday	No
14	Sunset	Weekend	No
15	Midday	Weekday	No

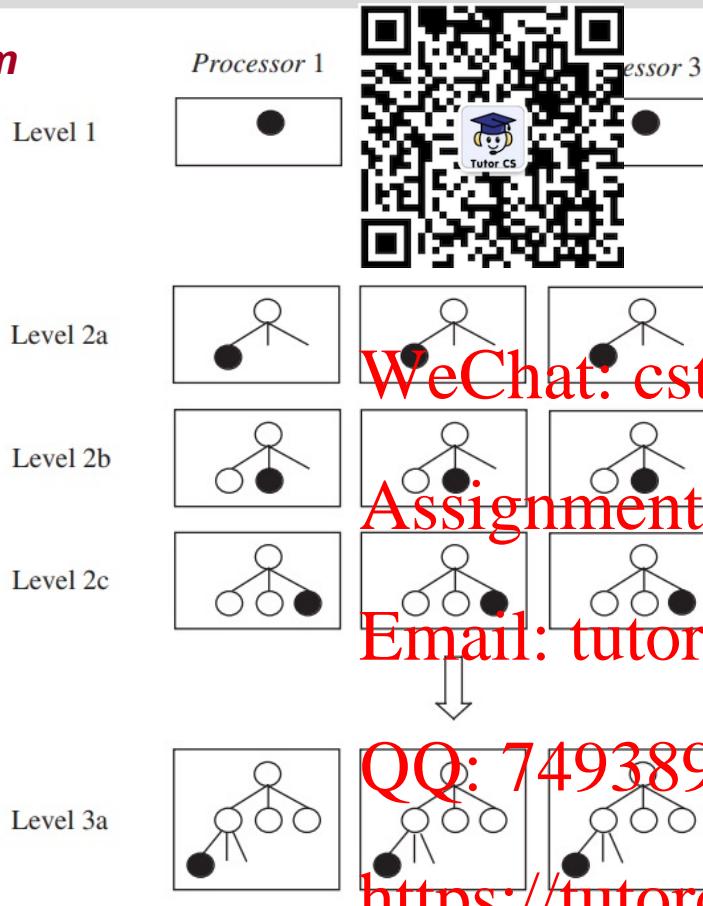
Partition 1

<https://tutorcs.com>

Partition 2

Parallel Classification: Decision Tree

Data parallelism



Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Figure 17.17 Data parallelism of parallel decision tree construction

Parallel Classification: Decision Tree

Level 1 (Root Node):

Processor 1

Rec#	Weather	Temperature	Target Class
1			
2			
...			
15			

Data parallel



Locally calculate the information gain values for: *Weather* and *Temperature*

Global information sharing stage:

- Share target class counts to calculate dataset entropy value
- Exchange dataset entropy value to determine splitting attribute (e.g. Time attribute is decided to be the splitting attribute)
- Distribute selected records# to all processor for the next phase (e.g. records 3, 6, 8, 9, 10, 15 for Time Midday, and records 1, 2, 5, 14 for Time Sunset)

Email: tutorcs@163.com

QQ: 749389476

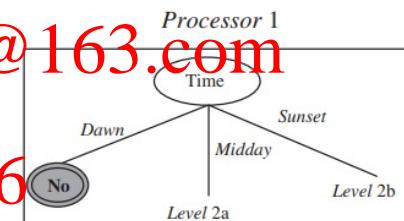
<https://tutorcs.com>

WeChat: cstutorcs

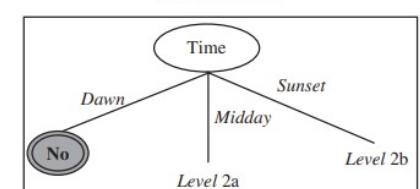
Locally calculate the information gain values for: *Time* and *Day*

Assignment Project Exam Help

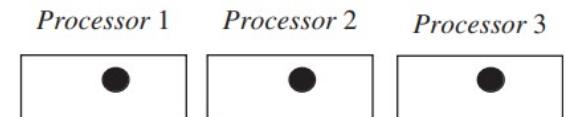
Decision tree for Level 1:



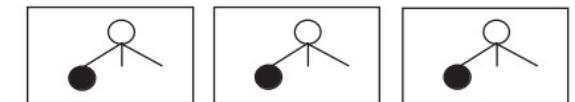
Processor 2



Level 1



Level 2a



Level 2b

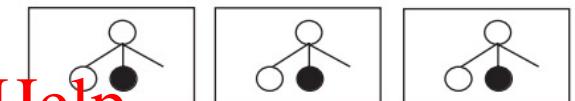


Figure 17.18 Data parallelism in decision tree

Parallel Classification: Decision Tree

Data parallelism

Level 2a:

Processor 1

Rec#	Weather	Temperature	TargetClass
3			
6			
8			
9			
10			
15			



WeChat: cstutorcs

Locally calculate the information gain values for: Weather and Temperature

Locally calculate the information gain values for: Day

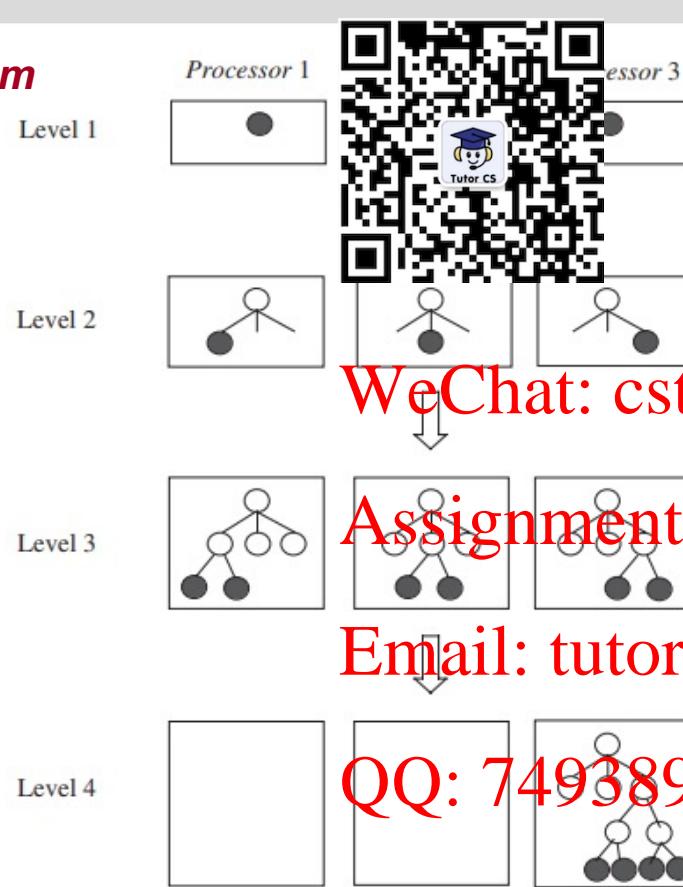
Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Parallel Classification: Decision Tree

Result parallelism



Rec#	Weather	Temperature	Time	Day	Jog (Target Class)
1	Fine	Mild	Sunset	Weekend	Yes
2	Fine	Hot	Sunset	Weekday	Yes
3	Shower	Mild	Midday	Weekday	No
4	Thunderstorm	Cool	Dawn	Weekend	No
5	Shower	Hot	Sunset	Weekday	Yes
6	Fine	Hot	Midday	Weekday	No
7	Fine	Cool	Dawn	Weekend	No
8	Thunderstorm	Cool	Midday	Weekday	No
9	Fine	Cool	Midday	Weekday	Yes
10	Fine	Mild	Midday	Weekday	Yes
11	Shower	Hot	Dawn	Weekend	No
12	Shower	Mild	Dawn	Weekday	No
13	Fine	Cool	Dawn	Weekday	No
14	Thunderstorm	Mild	Sunset	Weekend	No
15	Thunderstorm	Hot	Midday	Weekday	No

Figure 17.11. Training dataset

Figure 17.20 Result parallelism of parallel decision tree construction

Parallel Classification: Decision Tree

Result parallelism

Horizontal Data Partitioning:

Processor 1

Rec#	Weather	Temp	Time	Day	Target Class	Processor 2	Temp	Time	Day	Target Class
1					9					
2					10					
...					15					
8										



WeChat: cstutorcs

Level 1 (Root Node):

- Count target class on each partition
- Perform intra-node parallelism the same as for data parallelism to share target class counts to calculate dataset entropy value, exchange dataset entropy value to determine splitting attribute, and distribute selected records# to all other processors for the next phase)

QQ: 749389476

<https://tutorcs.com>

Parallel Classification: Decision Tree

Result parallelism

Jog	
Yes	No
5	10

Time	Dawn	5	
	Midday	6	
	Sunset	3 1 4	15

		Jog	
		Yes	No
Day	Weekend	4	6
	Weekday	1	4
			5
			15

WeChat: cstutorcs

		Jog	
		Yes	No
Weather	Fine	4	3 7
	Shower	1	3 4
	Thunderstorm	0	4 4 15

Assignment Project Exam Help

Email: tutorcs@163.com

Temperature

Hot

		Jog	
		Yes	No
Temperature	Hot	2	3 5
	Mild	3	2 5
	Cool	1	4 5
			15

QQ: 749389476

<https://tutorcs.com>

Parallel Classification: Decision Tree

Result parallelism

Level 2:

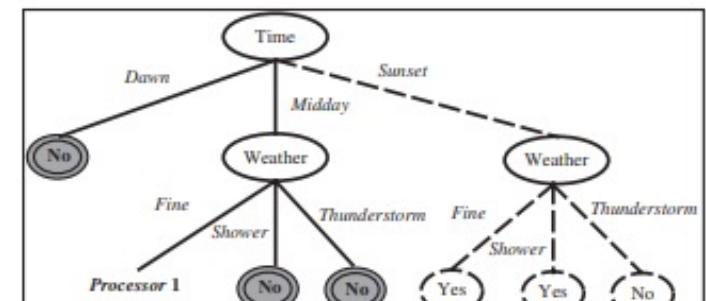
Processor 1							Processor 2						
Rec#	Weather	Temp	Time	Day	Target Class		Temp	Time	Day	Target Class			
3						1							
6						2							
8						3							
9						4							
10						14							
15													



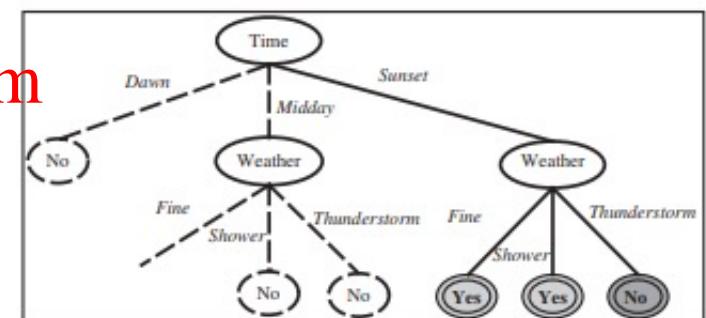
WeChat: cstutorcs

Result decision tree for Level 2:

Processor 1



Processor 2



Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Parallel Classification: Decision Tree

Level 3:

Processor 1

Rec#	Weather	Temp	Time	Day	Target Class
6					
9					
10					



Processor 2

Temp	Time	Day	Target Class

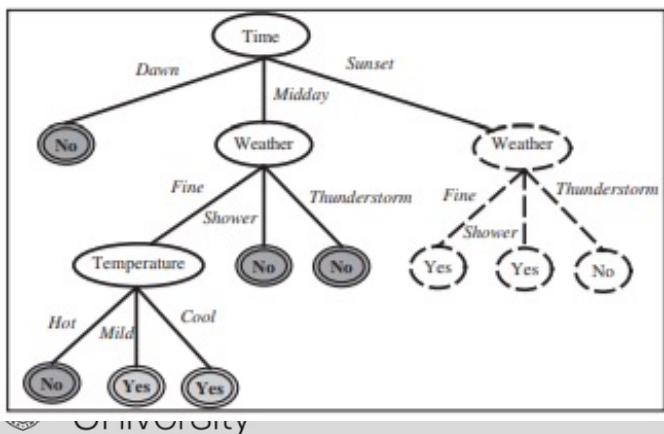


WeChat: cstutorcs

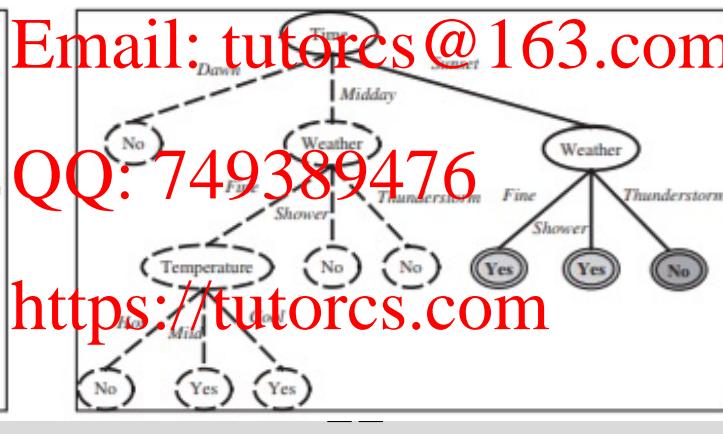
Global information sharing stage:... as like in Level 2 ...

Result decision tree for Level 3:

Processor 1



Processor 2

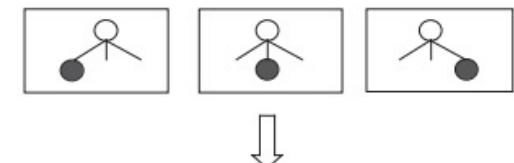


Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Level 2



Level 3

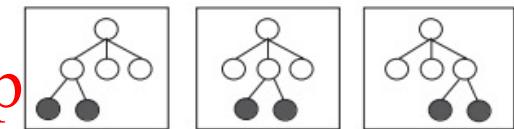


Figure 17.20 Result parallelism of parallel decision tree construction

See you next week!! 程序代写代做 CS编程辅导



Questions???

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>