Français (https://moodle.epfl.ch/mod/quiz/review.php?attempt=27053&lang=fr)
English (https://moodle.epfl.ch/mod/quiz/review.php?attempt=27053&lang=en)

Started on	Thursday, 26 April 2018, 10:49
State	Finished
Completed on	Thursday, 26 April 2018, 11:03
Time taken	13 mins 35 secs
Grade	**6.00** out of 8.00 (**75**%)

Question 1

Correct

Mark 1.00 out of 1.00

Flag question

In an FP tree, the leaf nodes are the ones with:

Select one:

- a. **Least in the alphabetical order**
- b. **None of the other options.**
- c. **Lowest confidence**
- d. **Lowest support** ✓

The correct answer is: **Lowest support**

Question 2

Incorrect

Mark 0.00 out of 1.00

Flag question

Suppose that q is density reachable from p. The chain of points that ensure this relationship are {t,u,g,r} Which one is FALSE?

Select one:

- a. **q has to be a border point**
- b. **{t,u,g,r} have to be all core points.** ✗ true ! They all have to be core points
- c. **p has to be a core point** false ! p can be a border point
- d. **p and q will also be density-connected**

The correct answer is: **q has to be a border point**

Question 3

Correct

Mark 1.00 out of 1.00

Flag question

Suppose that an item in a leaf node N exists in every path. Which one is correct?

Select one:

- a. **For every node P that is a parent of N in the fp tree, confidence(P->N) = 1**
- b. **N's minimum possible support is equal to the number of paths.** ✓

c. **N co-occurs with its prefix in every transaction.**

d. **The item N exists in every candidate set.**

The correct answer is: **N's minimum possible support is equal to the number of paths.**

Question 4

Correct

Mark 1.00 out of 1.00

Flag question

Fundamentally, why clustering is considered an unsupervised machine learning technique?

Select one:

a. **Number of clusters are not known.**

b. **The class labels are not known.** ✓

c. **The features are not known.**

d. **The clusters can be different with different initial parameters.**

The correct answer is: **The class labels are not known.**

Question 5

Correct

Mark 1.00 out of 1.00

Flag question

What is a correct pruning strategy for decision tree induction?

Select one:

a. **Remove attributes with lowest information gain.**

b. **Choose the model that maximizes L(M) + L(M|D)**

c. **Stop partitioning a node when either positive or negative samples dominate the samples of the other class.** ✓

d. **Apply Maximum Description Length principle**

The correct answer is: **Stop partitioning a node when either positive or negative samples dominate the samples of the other class.**

Question 6

Incorrect

Mark 0.00 out of 1.00

Flag question

When using bootstrapping in Random Forests, the number of different data items used to construct a single tree is:

Select one:

a. **Depends on the outcome of the sampling process, and can be both smaller or larger than the training set** ✗

b. **Of order square root of the size of the training set with high probability**

c. **Smaller than the size of the training data set with high probability**

d. **The same as the size of the training data set**

The correct answer is: **Smaller than the size of the training data set with high probability**

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

A1	P	N
a	7	0
b	1	4
A2	P	N
x	5	1
y	3	3

Given the distribution of positive and negative samples for attributes A1 and A2, which is the best attribute for splitting?

Select one:

a. **A2**

b. **A1** ✓

c. **They are the same**

d. **There is not enough information to answer the question**

The correct answer is: **A1**

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

Which of the following is true for a density based cluster C?

Select one:

a. **Any two points in C must be density connected. Each point belongs to one, and only one cluster**

b. **Any two points in C must be density reachable. Each point belongs to one, and only one cluster**

c. **Any two points in C must be density connected. Border points may belong to more than one cluster** ✓

d. **Any two points in C must be density reachable. Border points may belong to more than one cluster**

The correct answer is: **Any two points in C must be density connected. Border points may belong to more than one cluster**

Finish review (https://moodle.epfl.ch/mod/quiz/view.php?id=983960)