Started on	Thursday, 17 May 2018, 10:47
State	Finished
Completed on	Thursday, 17 May 2018, 11:05
Time taken	17 mins 54 secs
Grade	**6.00** out of 8.00 (**75**%)

Question 1

Correct

Mark 1.00 out of 1.00

Flag question

Which of the following is **correct** regarding *Crowdsourcing*?

Select one:

- a. *Random Spammers* give always the same answer for every question
- b. *Honey Pot* discovers all the types of spammers but not the *sloppy workers*
- c. It is applicable only for binary classification problems
- d. The output of *Majority Decision* can be equal to the one of *Expectation-Maximization* ✓

The correct answer is: The output of *Majority Decision* can be equal to the one of *Expectation-Maximization*

Question 2

Incorrect

Mark 0.00 out of 1.00

Flag question

Which of the following is **correct** regarding *Louvain* algorithm?

Select one:

- a. If n cliques of the same order are connected cyclically with n-1 edges, then the algorithm will always detect the same communities, independently of the starting point
- b. It creates a hierarchy of communities with a common root ✗
- c. *Clique* is the only topology of nodes where the algorithm detects the same communities, independently of the starting point
- d. *Modularity* is always maximal for the communities found at the top level of the community hierarchy

The correct answer is: If n cliques of the same order are connected cyclically with n-1 edges, then the algorithm will always detect the same communities, independently of the starting point

Question 3

Correct

Mark 1.00 out of 1.00

Flag question

Which is an appropriate method for fighting skewed distributions of class labels in classification?

Select one:

- a. Generate artificial data points for the most frequent classes
- b. Construct the validation set such that the class label distribution approximately matches the global distribution of the class labels ✓
- c. Include an over-proportional number of samples from the larger class
- d. Use leave-one-out cross validation

The correct answer is: Construct the validation set such that the class label distribution approximately matches the global distribution of the class labels

Question 4

Correct

Mark 1.00 out of 1.00

Flag question

		Class	
		Fraud	¬Fraud
Classified	Fraud	20	20
	¬Fraud	10	60

Considering the results of this fraud classifier, which of the following is **correct**?

Select one:

- a. The classifier has a precision of 50% and a recall of 75%
- b. The classifier has a precision of 75% and a recall of 50%
- c. The classifier has a precision of 66.6% and a recall of 75%
- d. The classifier has a precision of 50% and a recall of 66.6% ✓

The correct answer is: The classifier has a precision of 50% and a recall of 66.6%

Question 5

Incorrect

Mark 0.00 out of 1.00

Which of the following is **correct** regarding prediction models?

Select one:

- a. Training error being less than test error means overfitting ✗
- b. Simple models have lower bias than complex models

Flag question

c. Training error being less than test error means underfitting

d. Complex models tend to overfit, unless we feed them with more data

The correct answer is: Complex models tend to overfit, unless we feed them with more data

Question 6

Correct

Mark 1.00 out of 1.00

Flag question

Which of the following is **correct** regarding community detection?

Select one:

a. The *Girvan-Newman* algorithm attempts to maximize the overall betweenness measure of a community graph

b. High modularity of a community indicates a large difference between the number of edges of the community and the number of edges of a null model ✓

c. High betweenness of an edge indicates that the communities are well connected by that edge

d. The *Louvain* algorithm attempts to minimize the overall modularity measure of a community graph

The correct answer is: High modularity of a community indicates a large difference between the number of edges of the community and the number of edges of a null model

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

In the χ^2 statistics for a binary feature, we obtain $P(\chi^2 \mid DF = 1) > 0.05$. This means in this case, it is assumed:

Select one:

a. That the class labels depends on the feature

b. That the class label correlates with the feature

c. None of the above

d. That the class label is independent of the feature ✓

The correct answer is: That the class label is independent of the feature

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

In *User-Based Collaborative Filtering*, which of the following is **correct**, assuming that all the ratings are positive?

Select one:

a. If the ratings of two users have both variance equal to 0, then their *Cosine Similarity* is maximized ✓

b. *Pearson Correlation Coefficient* and *Cosine Similarity* have the same value range, but can return different similarity ranking for the users

c. *Pearson Correlation Coefficient* and *Cosine Similarity* have different value range, but return the same similarity ranking for the users

d. If the variance of the ratings of one of the users is 0, then their *Cosine Similarity* is not computable

The correct answer is: If the ratings of two users have both variance equal to 0, then their *Cosine Similarity* is maximized

Finish review (https://moodle.epfl.ch/mod/quiz/view.php?id=986296)