

Started on Thursday, 17 May 2018, 10:47

State Finished

Completed on Thursday, 17 May 2018, 11:05

Time taken 18 mins

Grade 7.00 out of 8.00 (88%)

Question 1

Correct

Mark 1.00 out of 1.00

Flag question

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

Select one:

- ☐ a. That the class label correlates with the feature
- ☐ b. None of the above
- ☐ c. That the class labels depends on the feature
- ☒ d. That the class label is independent of the feature ✓

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

Question 2

Correct

Mark 1.00 out of 1.00

Flag question

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

Select one:

- ☐ a. Training error being less than test error means overfitting
- ☒ b. Complex models tend to overfit, unless we feed them with more data ✓
- ☐ c. Training error being less than test error means underfitting
- ☐ d. Simple models have lower bias than complex models

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

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. Include an over-proportional number of samples from the larger class

- ☐ b. Generate artificial data points for the most frequent classes
- ☒ c. Construct the validation set such that the class label distribution approximately matches the global distribution of the class labels ✓
- ☐ d. Use leave-one-out cross validation

The correct answer is: Use leave-one-out cross validation

Question 4

Incorrect

Mark 0.00 out of 1.00

Flag question

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

Select one:

- ☒ a. It creates a hierarchy of communities with a common root
✗ false ! if the communities are disjoint
- ☐ b. *Clique* is the only topology of nodes where the algorithm detects the same communities, independently of the starting point false ! question in notes
- ☐ c. 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
- ☐ d. *Modularity* is always maximal for the communities found at the top level of the community hierarchy slides

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 5

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. *Pearson Correlation Coefficient* and *Cosine Similarity* have the same value range, but can return different similarity ranking for the users
- ☐ b. If the variance of the ratings of one of the users is 0, then their *Cosine Similarity* is not computable
- ☒ c. If the ratings of two users have both variance equal to 0, then their *Cosine Similarity* is maximized ✓
- ☐ d. *Pearson Correlation Coefficient* and *Cosine Similarity* have different value range, but return the same similarity ranking for the users

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

Question 6

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

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

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

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

Select one:

- ☐ a. The *Louvain* algorithm attempts to minimize the overall modularity measure of a community graph
- ☐ b. The *Girvan-Newman* algorithm attempts to maximize the overall betweenness measure of a community graph
- ☒ c. High betweenness of an edge indicates that the communities are well connected by that edge ✓
- ☐ d. 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

The correct answer is: High betweenness of an edge indicates that the communities are well connected by that edge

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

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

Select one:

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

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

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