

## 2019 Final - MCQ

1. Which one of the following about the Expectation-Maximization algorithm is FALSE?
  - a) In E step the labels change, in M step the weights of the workers change.
  - b) The label with the highest probability is assigned as the new label.
  - c) Assigning equal weights to workers initially decreases the convergence time. (CORRECT)
  - d) It distinguishes experts from normal workers.
2. Which of the following is TRUE?
  - a) Ontologies are used to directly map two schemas in order to overcome semantic heterogeneity.
  - b) Graph representation of an RDF statement facilitates exchange and storage.
  - c) RDF is a standardized model for encoding ontologies (CORRECT)
  - d) XML does not facilitate introducing new terms which are domain specific.
3. Regarding features engineering, which of the following is wrong:
  - a. Supervised discretization can merge any two intervals of the same variable. (CORRECT)
  - b. Classifiers can be sensitive to the absolute scale of the variables.
  - c. Features *filtering* consider single variables, whereas *wrapping* considers features combinations.
  - d. Standardisation can produce arbitrarily large values whereas scaling does not.
4. Given the graph  $1 \rightarrow 2$ ,  $1 \rightarrow 3$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 2$ , switching from Page Rank to Teleporting PageRank will have an influence on the value(s) of
  - a. All the nodes (CORRECT)
  - b. Node 1
  - c. Node 2 and 3
  - d. No nodes. The values will stay unchanged.
5. Which of the following is true:
  - a. Modularity is a measure of how communities are connected together
  - b. Agglomerative algorithms recursively decompose communities into sub-communities
  - c. Divisive algorithms are based on modularity
  - d. Girvan-Newman works by removing edges with the highest betweenness measure (CORRECT)
6. Which of the following is true:
  - a. The tf-idf weight is the ratio between tf and idf
  - b. The idf term decrease the impact of stop-words (CORRECT)
  - c. Frequent terms obtain low tf score
  - d. The tf term is computed over the whole document collection

7. Which of the following tasks would typically not be solved by clustering?

- a. Community detection in social networks.
- b. Discretization of continuous features.
- c. Spam detection in an email system (CORRECT)
- d. Detection of latent topics in a document collection

8. Which one is false about Label Propagation?

- a. The labels are inferred using the labels that are known apriori.
- b. It can be interpreted as a random walk model.
- c. Propagation of labels through high degree nodes are penalized by low abandoning probability. (CORRECT)
- d. Injection probability should be higher when labels are obtained from experts than by crowdworkers.