Iniziato	Monday, 9 January 2023, 09:55
Stato	Completato
Terminato	Monday, 9 January 2023, 11:19
Tempo impiegato	1 ora 23 min.
Valutazione	Non ancora valutato

Domanda **1**

Completo

Punteggio max.: 20,00

The task is described in this document.

The data file is here. The file with the feature names is here

Upload only your notebook, not the data. Please name your notebook according to the directions given in the document linked above.

chiara_angileri.ipynb

Domanda **2**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Which of the following is not a property of a *metric* distance function

Scegli un'alternativa:

- a. Boundedness ✔
- b. Positive definiteness
- c. Symmetry
- d. Triangle inequality

Risposta corretta.

La risposta corretta è: Boundedness

Domanda **3**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

- a. 0.5 ✔
- b. 0.3
- c. 0.2
- d. 0.1

Risposta corretta.

La risposta corretta è: 0.5

Domanda **4**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

What is the *single linkage?*

Scegli un'alternativa:

- a. A method to compute the distance between two sets of items, it can be used in hierarchical clustering ✔
- b. A method to compute the distance between two objects, it can be used in hierarchical clustering
- c. A method to compute the distance between two classes, it can be used in decision trees
- d. A method to compute the separation of the objects inside a cluster

Your answer is correct.

La risposta corretta è: A method to compute the distance between two sets of items, it can be used in hierarchical clustering

Domanda **5**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *recall* of a binary classifier?

Scegli un'alternativa:

a. (TP + TN) / (TP + FP + TN + FN)

b. TP / (TP + FN)

✔ This is also called *sensitivity*, or *hit rate*, which is the number of detected true positives divided by the total number of positives

c. TN / (TN + FP)

d. TP / (TP + FP)

Risposta corretta.

La risposta corretta è: TP / (TP + FN)

Domanda **6**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Why do we *prune* a decision tree?

Scegli un'alternativa:

- a. To eliminate parts of the tree where the decisions could be influenced by random effects ✔
- b. To eliminate parts of the tree where the decision could generate *underfitting*
- c. To eliminate attributes which could be influenced by random effects
- d. To eliminate rows of the dataset which could be influenced by random effects

Your answer is correct.

La risposta corretta è: To eliminate parts of the tree where the decisions could be influenced by random effects

Domanda **7**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

A Decision Tree is...

Scegli un'alternativa:

- a. A tree-structured plan of tests on single attributes to forecast the target ✔
- b. A tree-structured plan of tests on multiple attributes to forecast the target
- c. A tree-structured plan of tests on single attributes to forecast the cluster
- d. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node

Risposta corretta.

La risposta corretta è: A tree-structured plan of tests on single attributes to forecast the target

Domanda **8**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

When training a neural network, what is the *learning rate*?

- a. A multiplying factor of the correction to be applied to the connection weights ✔
- b. The speed of convergence to a stable solution during the learning process
- c. The slope of the activation function in a specific node
- d. The ratio between the size of the hidden layer and the input layer of the network

Your answer is correct.

La risposta corretta è: A multiplying factor of the correction to be applied to the connection weights

Domanda **9**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Which of the following is a strength of the clustering algorithm DBSCAN?

Scegli una o più alternative:

- a. Ability to find cluster with concavities ✔
- b. Ability to separate outliers from regular data ✔
- c. Very fast computation
- d. Requires to set the number of clusters as a parameter

Your answer is correct.

Le risposte corrette sono: Ability to find cluster with concavities, Ability to separate outliers from regular data

Domanda **10**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Which of the following characteristic of data can reduce the effectiveness of K-Means?

Scegli un'alternativa:

- a. Presence of outliers ✔
- b. All the variables are the same range of values
- c. All the variables have the same distribution of values
- d. Presence of values with high frequency

Your answer is correct.

La risposta corretta è: Presence of outliers

Domanda **11**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

Scegli una o più alternative:

- a. Reduce the minimum number of objects in the neighborhood ✔
- b. Increase the radius of the neighborhood ✔
- c. Decrease the radius of the neighborhood
- d. Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood

Risposta corretta.

Le risposte corrette sono: Reduce the minimum number of objects in the neighborhood, Increase the radius of the neighborhood

Domanda **12**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Which of the following statements regarding the discovery of association rules is true? (One or more)

Scegli una o più alternative:

- [x] a. The confidence of a rule can be computed starting from the supports of itemsets ✔
- [x] b. The support of an itemset is anti-monotonic with respect to the composition of the itemset ✔
- [ ] c. The confidence of an itemset is anti-monotonic with respect to the composition of the itemset
- [ ] d. The support of a rule can be computed given the confidence of the rule

Your answer is correct.

Le risposte corrette sono: The confidence of a rule can be computed starting from the supports of itemsets, The support of an itemset is anti-monotonic with respect to the composition of the itemset

Domanda **13**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Consider the transactional dataset below

ID	Items
1	A,B,C
2	A,B,D
3	B,D,E
4	C,D
5	A,C,D,E

Which is the *confidence* of the rule A,C ⇒ B?

Scegli un'alternativa:

- a. 50% ✔ 1 / 2
- b. 100%
- c. 40%
- d. 20%

Risposta corretta.

La risposta corretta è: 50%

Domanda **14**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

When is polynomial regression appropriate?

- a. When the relationship between the predicting variable and the target cannot be approximated as linear ✔
- b. When the target values are not linearly separable
- c. When there is more than one predicting attribute
- d. When it is necessary to project the data into a higher dimensional space

Your answer is correct.

La risposta corretta è: When the relationship between the predicting variable and the target cannot be approximated as linear

Domanda **15**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

Which is the purpose of discretisation?

Scegli un'alternativa:

- a. Reduce the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities ✔
- b. Reduce the number of distinct values in an attribute, in order to increase the efficiency of the computations
- c. Increase the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities
- d. Reduce the range of values of a numeric attribute, to make all the attributes more comparable

Risposta corretta.

La risposta corretta è: Reduce the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities

Domanda **16**

Risposta corretta

Punteggio ottenuto 0,67 su 0,67

In which part of the CRISP methodology we perform the **test design** activity?

- a. Modeling ✔
- b. Evaluation
- c. Business Understanding
- d. Data Understanding

Your answer is correct.

La risposta corretta è:
Modeling