← Lesson 2 Quiz
Quiz, 6 questions

1 point

1. The following real dataset contains information about two different flowers: Iris setosa and Iris versicolor.

Species	Sepal length	Sepal width	Petal length	Petal width
Iris setosa	4.9	3.0	1.4	0.2
Iris versicolor	5.6	2.5	3.9	1.1

What is the Manhattan distance between these two objects?

○ 2.5

|5.6-4.9|+|2.5-3|+|3.9-1.4|+|1.1-0.2| = 4.6

○ 7.8

○ 4.6

○ 2.8

1 point

2. The following real dataset contains two samples from the dataset for Prediction of Molecular Bioactivity for Drug Design – Binding to Thrombin, with sampled features. For each activity (F1, F2, ..., F10), the class value (0/1) indicates if the activity is active or inactive.

Cases	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
1	0	1	1	0	0	0	1	1	1	1
2	0	1	0	0	1	0	1	0	0	1

Assume all the activities are *symmetric* binary variables. What is the distance between case 1 and case 2?

○ 3/10

obs 2

symmetric = 4/10
asymmetric = 4/7

○ 4/7

		1	0
obs 1	1	3	3
	0	1	3

○ 4/10

○ 3/7

1
point

3. The following real world dataset contains two samples from Car Evaluation Database, which was derived from a simple hierarchical decision model originally developed for the demonstration of DEX (Bohanec, M., & Rajkovic, V. (1990). Expert system for decision making. Sistemica *1*(1), 145-157.). The model evaluates cars according to the following concept structure:

CAR	car acceptability
.PRICE	overall price
. . buying	buying price
. . maint	price of the maintenance
. TECH	technical characteristics
. . COMFORT	comfort
. . . doors	number of doors
. . . persons	capacity in terms of persons to carry
. . . lug_boot	the size of luggage boot
. . safety	estimated

The attribute values are as follows:

Attribute	Values (categorical)
buying	v-high, high, med, low
maint	v-high, high, med, low
doors	2, 3, 4, 5 - more
persons	2, 4, more
lug_boot	small, med, big
safety	low, med, high

Case	buying	maint	doors	persons	lug_boot	safety
Car 1	med	v-high	3	more	small	med
Car 2	high	v-high	4	4	big	med

To calculate the distance between objects with categorical attributes, we use a set of binary attributes to represent each categorical attribute. Assume all the binary attributes are **asymmetric**. What is the distance between Car 1 and Car 2?

○ 8/21

                                               asymmetric = diff/(all except both 0) = 4 (in orange) / 5 (orange+red)

○ 2/3

○ 8/10

○ 8/17

○ 1/3

4. Consider a two-dimensional space. Given a query point Q = (0.8, 0.6), which of the following is the closest to Q in terms of cosine similarity?

1 point

- ○ (6, 8)   **cosine similarity = d1 dot d2/ ‖d1‖*‖d2‖**

- ○ (-0.8, -0.6)   **(6,8) -> 9.6/ 10 = 0.96**
  (-.8, -.6) -> -1/1 = -1
  (.8, -.6) -> 0.28

- ○ (0.8, -0.6)   **(16,12) ->   20/20 = 1**

- ○ <mark>(16, 12)</mark>

5. Given the following two short texts with punctuation removed, calculate the cosine similarity between them based on the bag of words model.

1 point

Text1: one sees clearly only with the heart anything essential is invisible to the eyes

Text2: let my soul smile through my heart and my heart smile through my eyes that I may scatter rich smiles in sad hearts

- ○ 0.167

- ○ 0.201   **use python to calculate**

- ○ 0.009

- ○ <mark>0.117</mark>

6. With regard to the species of Iris versicolor, we have sampled data on the features of sepal length and sepal width, as follows.

1 point

pending

Feature	Sepal length	Petal length
Case 1	7.0	3.2
Case 2	6.4	3.2
Case 3	6.9	3.1
Case 4	5.5	2.3
Case 5	6.5	2.8

What is the correlation coefficient between sepal length and sepal width?

- ○ 0.342

- ○ 0.531

- ○ 0.804

◯ 0.882

☐ I, **Yuhui Chou**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Submit Quiz

👍 👎 ⚐