Haonan Hu

2/16/2021

2863545

In-Class Problems

In the correlation coefficient matrices below:

1. Which one(s) shows all of the features are independent? Why?
2. Which one(s) shows some, but not all, of the features are independent? Why?
3. Which one(s) show that features a and b are interdependent? Why?

Matrix A

	a	b	c
a	0.6	0.3	0.1
b	0.3	0.5	0.2
c	0.1	0.2	0.7

Matrix B

	a	b	c
a	1.0	0.0	0.0
b	0.0	1.0	0.0
c	0.0	0.0	1.0

Matrix C

	a	b	c
a	0.9	0.1	0.0
b	0.1	0.9	0.0
c	0.0	0.0	1.0

4. Assume you are working with a corpus of samples each with 3 features: a, b, and c, represented as (a,b,c) in feature space. If a training sample is located at (1,4,8), what is the Euclidean distance from the test sample (3,8,2) to the training sample?

For the kNN results shown at the bottom:

5. What is the class of the test sample if k = 3?
6. Why?

Training Sample (i)	Class	Euclidean Distance from Test Sample
1	Cat	3.9
2	Chicken	19.0
3	Dog	12.5
4	Chicken	15.7
5	Cat	7.5

1. Which one shows all the features are independent? Why?

   Matrix B, since B has all non-diagonal features are 0.

2. Which one shows some, but not all the features are independent? Why?

   Matrix C, since C has some non-diagonal features as 0, but not all.

3. Which one show that features a and b are interdependent? Why?

   Matrix A and C, because both <a, b>, <b, a> in both matrixes are non-zero, which mean they are interdependent

4.

Euclidean Distance (<3, 8, 2>, <1, 4, 8>) =

$$\sqrt[2]{(3-1)^2+(8-4)^2+(8-2)^2} = \sqrt[2]{56} \cong 7.48$$

5. What is the class of the test sample if k=3?

   Cat

6. Why?

First shortest distance is cat (3.9), next one is cat (7.5), third one is Dog (12.5), since 2/3 are cats, so the class of the test sample should be classified as cat.