Assignment - 2

Page
Date

* KNN (K-nearest neighbour)

(1) Following data are given:

	pepper	Ginger	chilly	Linked
A	True	True	True	false
B	True	false	false	True
C	false	True	True	false
D	false	True	false	True
E	True	false	false	True

→ If new data is given as
Q: pepper : false
ginger : true
chilly : true
classify the Q if K = 1 to 5

Sol^n:

→ By using hamming distance (both categories value of same dist will be zero else one)

	pepper	Ginger	Chilly	dist
A	1	1	1	1+0+0 = 1
B	1	0	0	1+1+1 = 3
C	0	1	1	0+0+0 = 0
D	0	1	0	0+0+1 = 1
E	1	0	0	1+1+1 = 3

- If k=1, Q will be classified as false

if k=2, Q will be classified as false (2 false)

If k=3, Q will be classified as false (2 false 1 true)

If k=4, Q will be classified either true or false (2 false, 2 true)

if k=5, Q will be true (2 false, 3 true)

(2) For the following data apply kNN

Sepal length	Sepal width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Setosa
8.4	3.4	Virginia
8.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginia
6.1	2.8	Versicolor
7.3	2.9	Virginia
6.0	2.7	Virginia
5.8	2.8	Versicolor
6.3	2.3	Versicolor
5.1	2.8	Versicolor
6.3	2.5	Versicolor
5.5	2.4	Versicolor

→ Q: Sepal length = 5.1
Sepal width = 3.1
species = ?

→ We will use Euclidean distance

Sepal length	Sepal width	Species	distance
5.8	3.7	setosa	$\sqrt{0.46}$
5.1	3.8	setosa	$\sqrt{0.53}$
4.2	3.0	virginica	$\sqrt{1.1}$
5.4	3.4	Setosa	$\sqrt{0.13}$
5.1	3.5	setosa	$\sqrt{0.5}$
5.4	3.9	setosa	$\sqrt{1.04}$
7.4	2.8	virginica	$\sqrt{3.04}$
6.1	2.8	virginica	$\sqrt{3.04}$
7.3	2.9	virginica	$\sqrt{4.3}$
6.0	2.7	versicolor	$\sqrt{0.37}$
5.8	2.8	virginica	$\sqrt{0.53}$
6.3	2.3	versicolor	$\sqrt{0.804}$
6.1	2.8	"	$\sqrt{0.53}$
6.3	2.8	"	$\sqrt{1.5}$
5.5	2.4	"	$\sqrt{0.57}$

If k = 5, min 5 distn from Q are 0.37, 0.44, 0.361, 0.61 and 0.67, in which close are three setosa, one virginica and one versicolor

Answer: species : setosa