Haonan Hu

3/30/2021

2863545

In-Class Problems

1. Calculate the Class Balanced Accuracy for the confusion matrix below.
2. Calculate the Balanced Accuracy for the confusion matrix below.

		True/Actual		
		Cat (🐱)	Fish (🐟)	Hen (🐓)
Predicted	Cat (🐱)	4	6	3
	Fish (🐟)	1	2	0
	Hen (🐓)	1	2	6

3. Calculate the distance in feature space (d) between the 2 iris dataset samples below.
4. Using SMOTE Oversampling with RN = 0.5, synthesize a sample of Iris-sesota between these 2 samples.

sample	sepal-length	sepal-width	petal-length	petal-width	class
real (r)	5.1	3.5	1.4	0.2	Iris-setosa
nearest neighbor (n)	4.9	3	1.4	0.2	Iris-setosa

5. Calculate the centroid in feature space of the 3 iris dataset samples below.

sepal-length	sepal-width	petal-length	petal-width	class
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa

1. Calculate the Class Balanced Accuracy for the confusion matrix below.

P(Cat) = 4 / (4+6+3) = 4/13

P(Fish) = 2 / 3

P(Hen) = 6 / 9 = 2 / 3

R(Cat) = 4 / 6 = 2 / 3

R(Fish) = 2 / 10 = 1 / 5

R(Hen) = 6 / 9 = 2 / 3

Min(P(Cat), R(Cat)) = 4 / 13

Min(P(Fish), R(Fish)) = 1 / 5

Min(P(Hen), R(Hen)) = 2 / 3

Class Balanced Accuracy = ((4/13) + (1/5) + (2/3)) / 3 = 39.15%

2. Calculate the Balanced Accuracy for the confusion matrix below.

R(Cat) = 4 / 6 = 2 / 3

R(Fish) = 2 / 10 = 1 / 5

R(Hen) = 6 / 9 = 2 / 3

Specificity(Cat) = (2 + 2 + 6) / ((2 + 2 + 6) + (6 + 3)) = 10 / 19

Specificity(Fish) = 14 / (14 + 1) = 14 / 15

Specificity(Hen) = (4 + 6 + 1 + 2) / ((4 + 6 + 1 + 2) + (1 + 2)) = 13 / 16

Avg(R(Cat), Specificity(Cat)) = 34 / 57

Avg(R(Fish), Specificity(Fish)) = 17 / 30

Avg(R(Hen), Specificity(Hen)) = 71 / 96

Balanced Accuracy = (34/57 + 17/30 + 71/96) / 3 = 63.32%

3. Calculate the distance in feature space (d) between the 2 iris dataset samples below.

d = $\sqrt[2]{(5.1-4.9)^2+(3.5-3)^2+(1.4-1.4)^2+(0.2-0.2)^2}$ = 0.539

4. Using SMOTE Oversampling with RN = 0.5, synthesize a sample of Iris-sesota between these 2 samples.

Synthetic sample = (5.1 + 0.5(4.9-5.1)), (3.5 **+** 0.5(3-3.5)), (1.4 + 0.5(1.4-1.4)), (0.2+0.5(0.2-0.2))

Sepal-length: 5

Sepal-width: 3.25

Petal-length: 1.4

Petal-width: 0.2

5. Calculate the centroid in feature space of the 3 iris dataset samples below.

[mean(5.1, 4.9, 4.7), mean(3.5, 3, 3.2), mean(1.4, 1.4, 1.3), mean(0.2, 0.2, 0.2)] = [4.9, 3.23, 1.36, 0.2]