Post-Lecture Quiz - Lecture 36: Classifiers

Total points 0/0

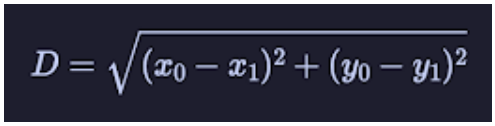
A quick conceptual review of what was gone over in lecture about classifiers. Note that there's no points associated with this Google Form — it is just here to see reinforce key points about the material!

Note: you can edit your responses after submission, but can only submit to this form once.

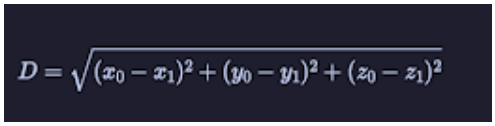
Helpful textbook chapter: 17.4

The respondent's email (**sean_villegas@berkeley.edu**) was recorded on submission of this form.

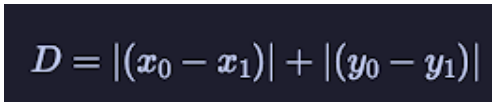
✓ What formula did we use in lecture to calculate the distance between points? Assume we have two attributes we are looking at. *



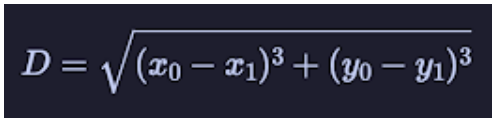
(●) Square root of the sum of squared differences for two terms ✓



( ) Square root of the sum of squared differences for three terms



( ) Sum of absolute differences



( ) Square root of the sum of cubed differences for two terms

Feedback

The 1st answer ("Euclidean Distance") is the correct answer. This is the formula you may see when calculating the straight line distance between two points. This choice assumes we are only looking at 2 attributes (i.e. has an x and y value only).

The 2nd answer is also Euclidean distance, but it looks at 3 attributes, and thus is incorrect (as we are looking just at 2 attributes as given in the question)

The 3rd answer uses absolute values ("Manhattan Distance"). This would be a valid distance metric, but it is not what is used in lecture.

The 4th answer choice cubes the differences rather than squaring it. Note that this allows for negative quantities (within the square root). However we need an even exponent to make sure the differences lead to a positive number (as we care about distance, not a direction).

Order the steps of finding the k Nearest neighbors! *

	1	2	3	4	Score	
Augment the training data table with a column containing all the distances	○	◉	○	○	0/0	✓
Sort the augmented table in increasing order of the distances	○	○	◉	○	0/0	✓
Find the distance between the example and each example in the training set	◉	○	○	○	0/0	✓
Take the top k rows of the sorted table	○	○	○	◉	0/0	✓

✗ We are creating a kNN classifier to classify a test point. The table below contains the labels of all the points in our training set and their distance to the test point, sorted in ascending order by distance. Using k = 5, what label would our classifier assign this test point? *

	Label	Distance
0	Green	0.0852
1	Red	0.1320
2	Green	0.1540
3	Red	0.4530
4	Red	0.6780
5	Green	0.6870
6	Red	0.9540
7	Green	1.4530
8	Green	1.5960
9	Red	2.3520

◉ Green ✗

○ Red

Correct answer

◉ Red

Feedback

The top 5 rows have the labels: Green, Red, Green, Red, and Red. Since there are 3 Red labels and 2 Green labels, the test point would be classified as Red.

✓ Let's use the same table as above and call it **colors**. Using .row() and .item(), how can we get the value 0.1320 (2nd row, 2nd column)? *

- [x] colors.row(1).item(1) ✓
- [x] colors.row(1).item("Distance") ✓
- [ ] colors.row(2).item(1)
- [ ] colors.row(2).item("Distance")

Feedback

The value 0.1320 is in the 2nd row and 2nd column. The second row corresponds with index 1 (.row(1)) and the 2nd column can be selected with either .item(1) or .item("Distance"). Note how we were able to use a label in .item() — this only allowed on row objects, not arrays.