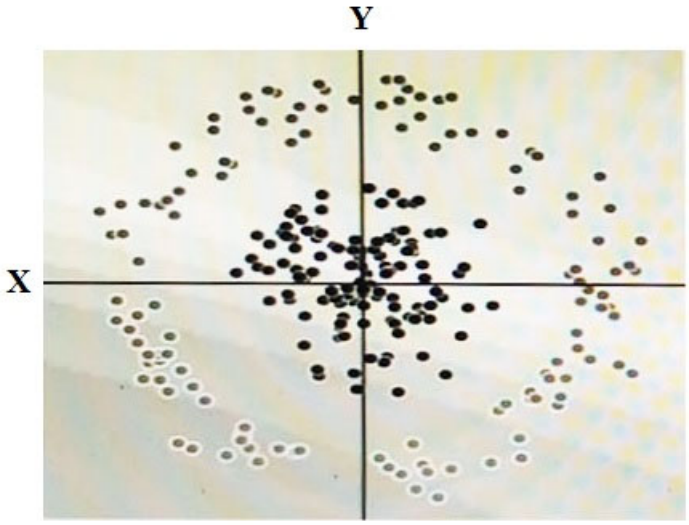
Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 63 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 63
Topic #: 1
[All Professional Data Engineer Questions]

You have some data, which is shown in the graphic below. The two dimensions are X and Y, and the shade of each dot represents what class it is. You want to classify this data accurately using a linear algorithm. To do this you need to add a synthetic feature. What should the value of that feature be?



A. X2+Y2

B. X2

C. Y2

D. cos(X)

Show Suggested Answer

by jvg637 at *March 16, 2020, 2:39 p.m.*

Comments

Type your comment...

Submit

jvg637 Highly Voted 5 years, 1 month ago
For fitting a linear classifier when the data is in a circle use A.
upvoted 41 times

[Removed] Highly Voted 5 years, 1 month ago
Answer: A
upvoted 15 times

SamuelTsch Most Recent 6 months, 2 weeks ago
Selected Answer: A
I think A should be x^2+y^2. We need a circle to classify the data.
upvoted 2 times

baimus 7 months, 2 weeks ago
Selected Answer: A
Just a note, they are using X2 and Y2 to mean Xsquared, and Ysquared. This is a circle in the form X2+Y2 = k, so for a given k will split that dataset nicely.
upvoted 1 times

Mathew106 1 year, 9 months ago
It's not obvious to me it is A.

As others said, cos(X) does ignore the Y value. But answer A does not seem good either. The differences seem minimal.

If you do A then you have the following issues. If you take elements in the bottom right or the top left of the circle, they will all have the same value, ZERO. Not only that, they will actually have the same value with the elements in the middle of the circle which are completely black. Moreover, elements on the extreme right and extreme right will have different values (-x_max and +x_max).

However, if you use a cos(x) then the elements in the beginning
upvoted 1 times

> Mathew106 1 year, 9 months ago
> Nevermind I did not understand that X2 and Y2 meant X^2 and Y2. Answer is A because that gives the distance from the circle. Circle radius = sqrt(X^2 + Y^2). So even though it's not a perfect answer, it makes sense.
> upvoted 3 times

samdhimal 2 years, 3 months ago
A. X2+Y2

The synthetic feature that should be added in this case is the squared value of the distance from the origin (0,0). This is equivalent to X2+Y2. By adding this feature, the classifier will be able to make more accurate predictions by taking into account the distance of each data point from the origin.

X2 and Y2 alone will not give enough information to classify the data because they do not take into account the relationship between X and Y.

D. cos(X) is not a suitable option because it does not take into account the Y coordinate.
upvoted 3 times

GCPpro 2 years, 3 months ago
A is the correct answer as graph of circle is x^2 + y^2
upvoted 2 times

desertlotus1211 2 years, 3 months ago

Answer is A:
The answer reflects 'x' to the 2nd power + 'y' the 2nd power.
I guess they can't use carots in the exam answers!

upvoted 1 times

AzureDP900 2 years, 4 months ago

A is right
Reference:
https://medium.com/@sachinkun21/using-a-linear-model-to-deal-with-nonlinear-dataset-c6ed0f7f3f51

upvoted 1 times

DipT 2 years, 4 months ago

Selected Answer: A

https://developers.google.com/machine-learning/crash-course/feature-crosses/video-lecture

upvoted 1 times

DGames 2 years, 4 months ago

Selected Answer: A

linear circle X2+Y2 https://www.stat.cmu.edu/~cshalizi/dm/20/lectures/08/lecture-08.html

upvoted 2 times

mvww11 2 years, 10 months ago

If the shape was a circle, it would be (x^2 + y ^2). But I think that a quadric curve will do a better job of separating the two classes, so it would be (x^2)

upvoted 2 times

gabrysave 2 years, 12 months ago

Answer: A.
X^2+Y^2 is the equation of a circle.

upvoted 1 times

diagniste 3 years ago

Selected Answer: A

C'est A

upvoted 2 times

Tanzu 3 years, 2 months ago

Selected Answer: A

only A is draw a circle

upvoted 1 times

sraakesh95 3 years, 3 months ago

Selected Answer: A

Equation of circle as represented in the question

upvoted 1 times

moumou 3 years, 3 months ago

F(x) as A B C will have always a positive values as result, for A will need a third dimenssion Z to represent data, only D:cos(x) can be presented as the shown classification. this is a math question

upvoted 1 times

NR22 3 years ago

A B C will only have positive values
imaginary numbers (i + j) : am I a joke to you?

upvoted 1 times

Load full discussion...