Daniel Campos November 14th , 2013

RCS: Campod2 RIN:660996361

Machine Learning CS 4100

I worked on these problems with Zoe Konrad☺

Problem Set 10

1. (100) Exercise 6.1
   1. Give two vectors with very high cosine similarity but very low distance similarity. Similarly give two vectors with very low cosine similarity but very high Euclidean distance similarity

Take vector A={1,1,1} and B={100,100,100}

Euclidean distance similarity = 171.47303 So low similarity

Cosine Distance Similarity=300/1.7320508\*173.20508= 300/299.999=1 so very high because its almost one

Oppositely take vector A={1,1,1} B={-4,-2,6}

Euclidean distance similarity = 7.681146 So high similarity

Cosine Distance Similarity=0/(1.71320508)\*(7.483314)

=0/12.9614=0 so very low cosine similarity

* 1. If we shift the origin of the coordinate system which of the two measurements will chance. How will tis affect your choice of features

I shifted my second set of vectors each by minus two so with the new origin A={-1,-1,-1} and B={-6,-4,4} with these shifts my Euclidean distance stayed the same at 7.681146 but my Cosine Distance shifted to 0.4200840. By changing the origin of the coordinate system our cosine distance will change but our Euclidean will not. This will make me choose features that the origin system will not shift, such as a range of always -1 to 1 instead of perhaps the amount of pixels for the digits. If I were ever able to get larger digits the amount of pixels would change so the coordinate system would. This change would cause the cosine distance to change but not the Euclidean.

1. (100) Exercise 6.2

Since if pi(x)>=.5 is +1 and -1 otherwise we have an equal distribution on x with half on each side. Because our system will tend towards either side we then just have to chose either he error for (x) or 1- (x)

For my hypothesis

1. (200) Problem 6.1
2. (200) Problem 6.4
3. (400) Problem 6.16