Ning

24 September 2019

Data Science

Class Summary

[OPENING]: Lots of topics regarding linear algebra as well as statistical concepts were introduced today.

[SUMMARY]: There was an opening discussion regarding the rank concept, explained through an example, where we take a point from a 2D space and project it into the 3D space. The rank of the resulting point is 2 still , even though we projected it into a 3D space because the point lies in a 2D plane, and has no third dimensionality (according to our function of projection) which makes the function not fully ranked. The trace of the matrix is the sum of the diagonal values. The trace values are invariant against basis changes.

[MEMO]: Should look up trace of a matrix.

[SUMMARY]: We went over an eigenspace concept, where you list out the eigen values in descending order, and notice that very quickly they scale to approaching zero. Therefore we do a sort of truncating and approximate the value of the matrix, and the goal here is to move to a lower rank matrix. This will be useful in PCA.

[MEMO]: We should look up the “Power Method”, “Gram-Schmidt Orthogonalization”, “Singular Value Decomposition” specifics.

[SUMMARY]: When we evaluate data points (i.e: in a 2D space) we need a concept of topology, neighborhood, all derived from a notion of “DISTANCE.” Nearest neighbors is a concept that is useful to talk about regarding data points. KNN is the K nearest neighbors, whereas is the nearest neighbors within sigma distance.

[MEMO]: We should look up the concept of a voronoi diagram, it’s based around points and distance.

[SUMMARY]: In conditional probability, we can say that we are altering the universe (1) to be the P(B) when we calculate P(A|B) = . You can technically interpret every probability P(X) as a conditional probability of P(X| everything not X). Baye’s Theorem also makes sense to be derived since you can see that is equal to .

[SUMMARY]: An informal definition of a random variable is a variable that can take certain values with a certain probability. Expectation and variance are terms that refer to the expected value of a variable, and how it differed in observation. Entropy is the measure of randomness. A *covariance matrix* has variance values on the diagonal since variance is just covariance of a variable with itself.

[SUMMARY]: If you have a “random” or uneven distribution, it is hard to sample from it “randomly” since the function inverse is hard to calculate. So therefore we need to estimate. This is the reason for estimation theory. We were introduced to the Weak Law of Large numbers, which is the idea that the probability as well as magnitude of your error shrinks as you increase the sample size. Central Limit Theorem suggests that the error you make when you increase the sample is modeled by a gaussian distribution.

[MEMO]: We should spend a lot more time fleshing out the Central Limit Theorem, it seems like it will be important.

[SUMMARY]: The curse of dimensionality: according to our sampling laws, we know that increasing the dimension will require us to increase the number of sampling in order to get the same estimation quality. However, sometimes increasing the sample size is costly or impossible (i.e: mice, cancer patients).

[MEMO]: We looked briefly at an example of high dimensionality modeled by a cube and sphere, where the more the dimensions increased, the more likely the points of relevance will propagate into the corners of the cube away from the sphere, making it SIGNIFICANTLY harder for us to find meaningful values (or any values for that matter). We should also look into this concept, it seems important to get a better intuition of high dimensionality.