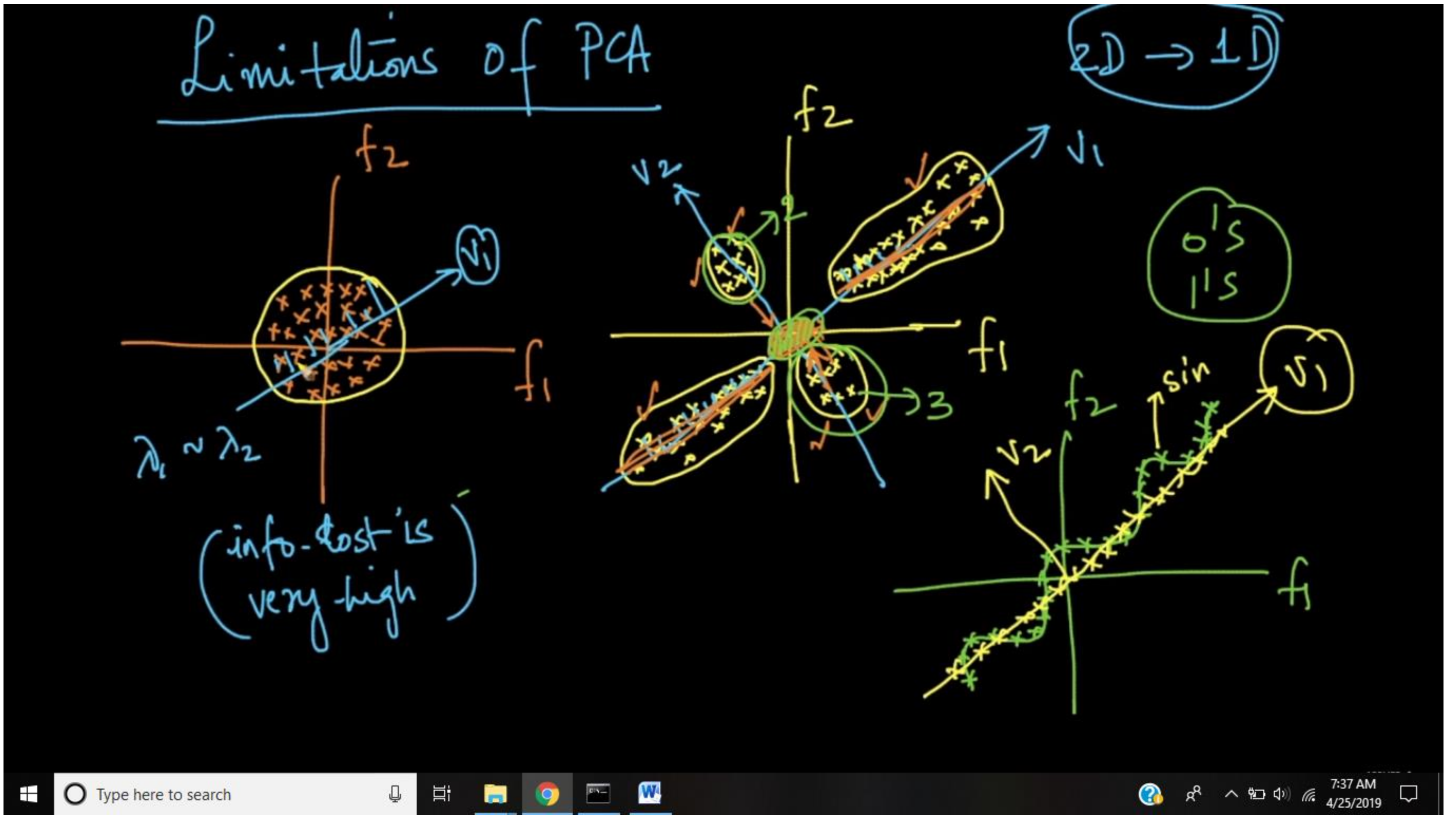

Let' take example where we are converting 2-D to 1-D, here are following observations:

1. For first figure where data is present in a form of circle, and for representing them we are just placing a vector within them, so all points will now projected to vector v1, and therefore there will very high lost of information.
   This means that if the data spread in circlular shape in 2d or hypersphere in nd in that case Lambda1 is roughly equal to lambda2 means roughly equal amount of spread of data on both the Vectors so if you transform it from 2d to 1d almost half of information is lost and you have left with only 50% of information/variance with you implies Information lost is very high.

2. In second figure there are 4 clusters of information, it's fine for 2 clusters that are already on v1, but 2 small cluster that are distinguishing themselves in 2-D, will now be projected to v1(since variance Is maximum on v1) in the same variance and hence will not be able to distinguish themselves, that means we are losing classification property here.

3. In third figure there is wave like data, but if we draw v1, then all will be projected on v1, with uniform variance, and there we lost this wave information.

   While we can project the points onto a 1-D plane/line which is 45-degrees, we are obtaining a bunch of points on a line. We have completely lost the actual structure, the sine-wave, of the original data(which is very useful when dealing with time-series data or fourier transform). The objective of any visualization is to capture as much structure of the original high-dim points as possible in the low-dim space.

basically, by projecting all the points in to 1D (in the direction of V1 where variance is maximum) and visualizing it, we will lose the sinusoidal shape.this is one of the main limitation of PCA.if-else logic will not be able to preserve the shape.

Comments:

1. why is capture the shape of data imp?

   Because the objective of any visualization is to capture as much structure of the original high-dim points as possible in the low-dim space as we dont want to completely lost the structure of actual data

2. As in 3rd case it seems PCA is not good for non-linear data, is it right?

   Its not prefer to apply pca as pca loses some of the interesting features just like in case of sine wave(which is very useful when dealing with time-series data or fourier transform). This means that if you have some of the variables in your dataset that are linearly correlated, PCA can find directions that represents your data, but if the data is not linearly correlated (f.e. in spiral, where x=t*cos(t) and y =t*sin(t)), PCA is not enough.

3. As i understand, whole Point of PCA is to bring down the dimensions so that it is computationally feasible/visually interpretable, while retaining as much information as possible in terms of variance. My question is why are we focused only on Variance, is it the only parameter of information. Why not other statistics or for that matter the original distribution of data? Could you please shed some light on this.

   Lots of variability usually indicates signal, whereas little variability usually indicates noise. Thus, the more variability there is in a particular direction is, theoretically, indicative of something important we want to detect. This is because covariance matrix accounts for variability in the dataset, and variability of the dataset is a way to summarize how much information we have in the data .

4. so data points should be correlated with each other before PCA?
   and after they will be uncorrelated as we are selecting eigen vectors with max spread which are orthogonal to each other

   Yes. After PCA, we don't see any feature correlated to any other feature and every new feature will be orthogonal to all other features.

5. How do we spot these PCA limitations if we are in higher dimensional spaces and we cannot see or envision these obvious limitations to start with?

   One way to understand this in high-dimensions would be to look at the preservation of variance when we project from higher-dimensions to lower-dimensions. If most of the variance from d-dim is preserved to d'-dim where d' < d, then, we can be very sure that most of the information is captured in lower dimensions. Hence, these limitations-cases would be less likely occur.