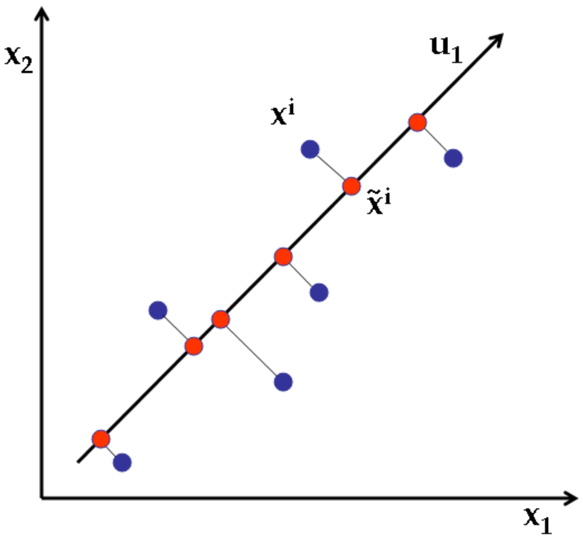
What happens when there are too many features?

Dimensionality Reduction!



Real-world data, such as speech signals, digital photographs, or fMRI scans, usually has a high dimensionality. In order to handle such real-world data adequately, its dimensionality needs to be reduced. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data.

Dimensionality reduction is important in many domains, since it mitigates the curse of dimensionality and other undesired properties of high-dimensional spaces. As a result, dimensionality reduction facilitates, among others, classification, visualization, and compression of high-dimensional data.

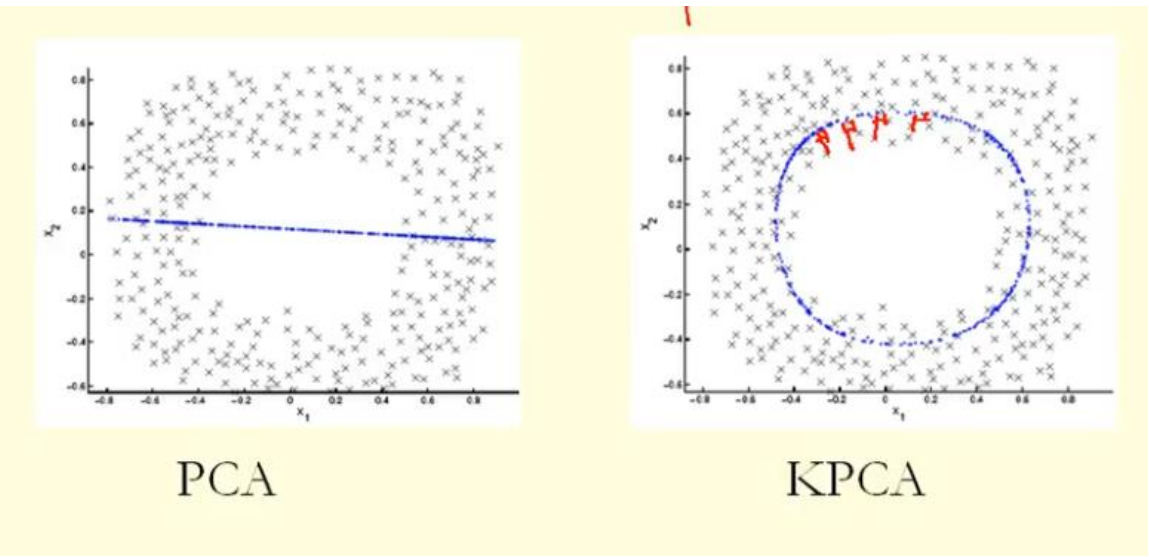
PCA / Kosambi-Karhunen-Loeve Transform –

It is a statistical procedure that uses an orthogonal transformation to convert a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the smaller number of original variables. In the field of Signal Processing, it is called the KLT transform.

The Algorithm –

1) Organize the dataset such that for n data-points of p dimensions, we form a (n*p) matrix
2) Normalize the dataset, by making it zero mean and unit SD
3) Compute the co-variance matrix by $S = X^TX$
4) Find the eigenvectors and the eigenvalues, and rearrange the eigenvectors according to decreasing value of the eigenvalues
5) Take the first k columns of the eigenvector matrix, and this is the required transform for our data

While this algorithm is a linear mapping, Kernels can be used which make the mapping non-linear. This is then called Kernel PCA. Details of this are beyond the scope of this presentation.



As seen the data is projected onto a highly non linear subspace using kernels.