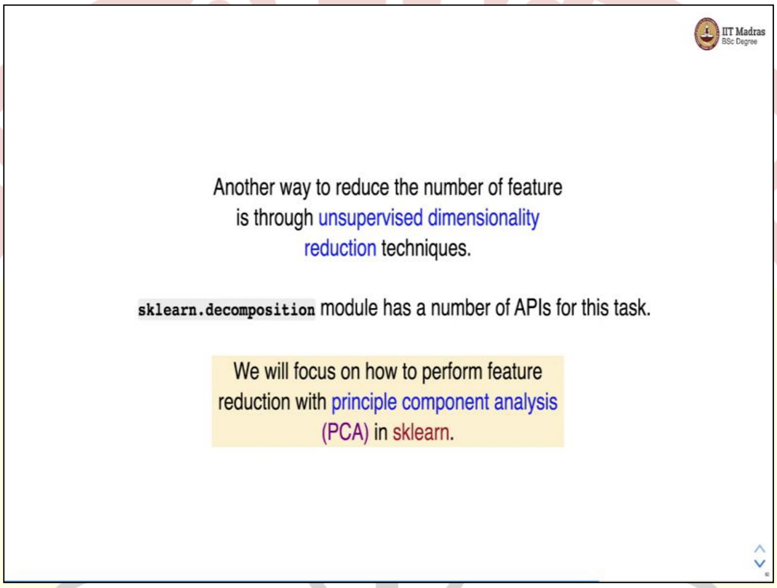
# IIT Madras

ONLINE DEGREE

**Machine Learning Practice**
**Online Degree Programme**
**B. Sc in Programming and Data Science**
**Diploma Level**
**Dr. Ashish Tendulkar**
**Indian Institute of Technology – Madras**

# Dimensionality Reduction by PCA

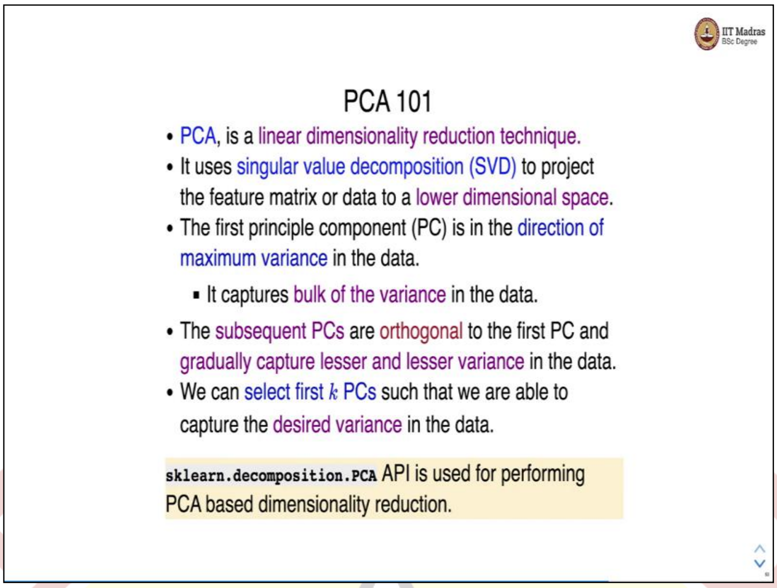
**(Refer Slide Time: 00:22)**




Namaste welcome to the next video of the machine learning practice course. In this video we will study dimensionality reduction. So, far we looked at wrapper based and filter based feature selection methods. The object of the feature selection method is to reduce the number of features. So, that it becomes more efficient to train the model. Another way to reduce the number of features is through unsupervised dimensionality reduction techniques.

Sklearn. decomposition module has a number of APIs for this task. We will focus on how to perform feature reduction with a specific technique called principal component analysis and we will focus on how to perform PCA in Sklearn.

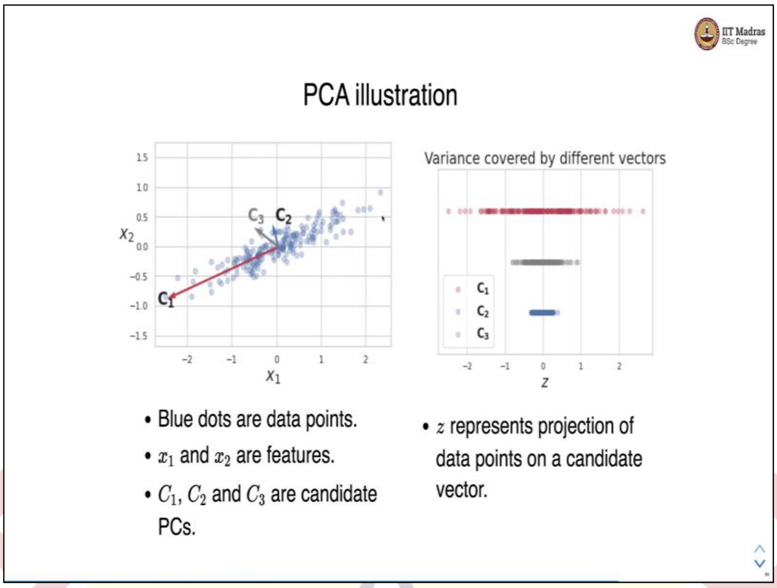
**(Refer Slide Time: 01:02)**




So, let us have a short tour of principal component analysis. The principal component analysis will be covered in greater detail from a theoretical perspective in later weeks of machine learning techniques. So, principal component analysis is a linear dimensionality reduction technique it uses singular value decomposition or SVD to project the feature metrics or data to lower dimension space. The first principal component is in the direction of maximum variance in the data in other words the first principal component captures the bulk of the variance in the data.

The subsequent principal components are also going into the first principal component and they gradually capture lesser and lesser variance in the data. We select for escape principal component such that we are able to capture the design variance in data. Sklearn under the composition of PCA API is used for performing principal component analysis based on dimension reduction.

**(Refer Slide Time: 02:11)**

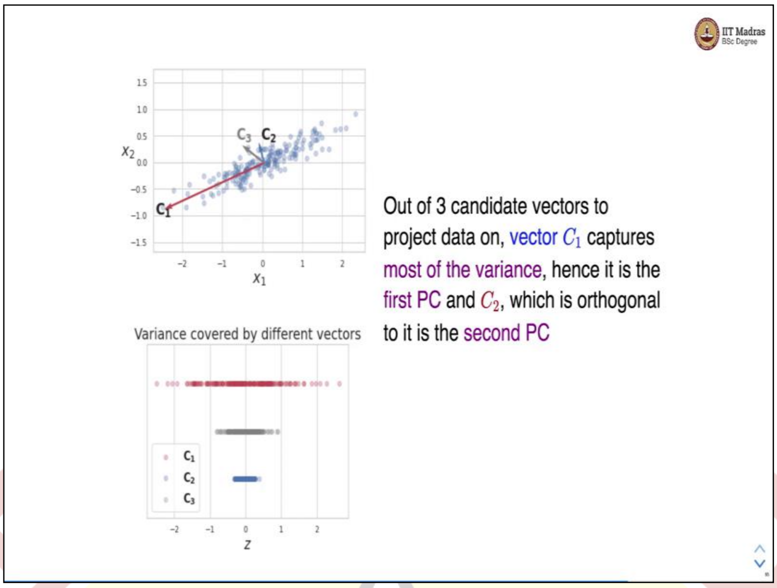



Let us illustrate PCA to run example so in this screen you see a training data with two features x1 and x2 and the blue points were here other training samples in the feature space mounted by x1 and x2. Now we are trying to evaluate three vectors for being the first principal components these three vectors are C1, C2, and C3. So, what we do is we try to find out we try to capture the variance covered by these 3 vectors for doing that what we do is the first project the original data points on these vectors and we find out what is a variance captured by each of the vectors.

So, that is the projection of the original data points to the principle to the candidate principal component vectors. So, the red points over here are the projection of the data points on the first principal component. The grace points are projections of the original data points on the second principle on the third principal component which is in green.

And then we have the bluish kind of projection which is the projection of the original data point on vector C2. So, you can easily see that the maximum variance or the bulk of the variance is captured by the vector C1.

**(Refer Slide Time: 04:01)**

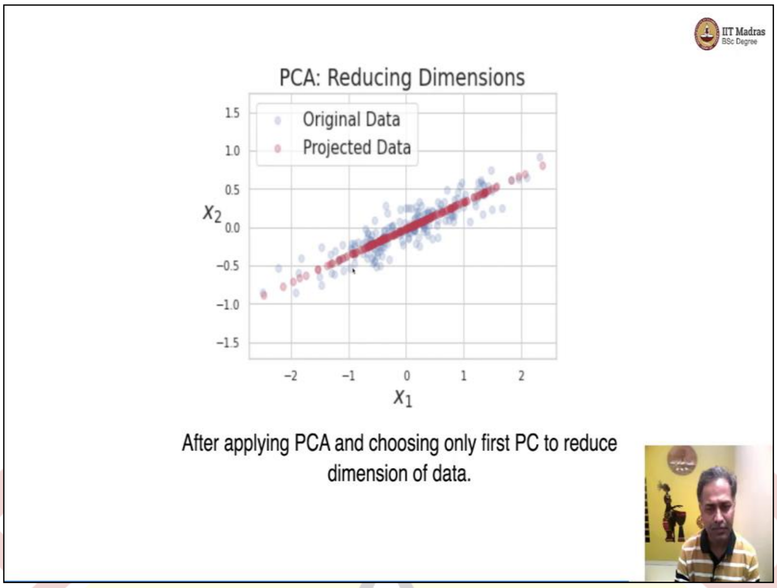



So, out of these three candidate principal components the vector C1 captures most of the variance in the data. I mean we make C1 our first principal component. Like; even though C3 captures the lesser variance than C1 but better than C2 we do not select C3 as a second principal component because the subsequent principal components have to be orthogonal to the first principal component.

Hence we select C2 as a second principal component because, among different vectors that are orthogonal to C1, C2 captures most of the variance in the data, and C2 becomes our second principal component and when we capture when we select these orthogonal vectors as principal components we make sure that these new feature vectors that are derived from the original feature vectors are independent of each other.

So, there is no relationship between these new features and that kind of thing helps us in reducing the number of features.

**(Refer Slide Time: 05:22)**




Now what we do is if you want to reduce the number of features from two to one then obviously the first principle component is our best brand because the first principal component covers the maximum variance in the data and what do you see here in red is the projection of the viewpoints on the first principle component. So, you can see that it has covered maximum variance in the data and that variance is even more than the original features x1 and x2.

So, that is its principal component analysis. So, we will be using principal component analysis for the unsupervised way of dimension reduction or the future reduction instead of more technical details about the principal component analysis in the later weeks of the machine learning techniques course.