

1.59.18

Module 4

Summary

In this session we will go through the individual steps of PCA

Before we do this let's make two statements:

① When we derive PCA, we make the assumption that our data is centered, which means it has mean 0

This assumption is not necessarily derived from PCA, right PCA and would have come to same result,

but subtracting the mean from the data can avoid ^{numerical} difficulties

Assume the values of our data is centered around 10^8 , then

Computing the data covariance matrix, requires us to multiply huge numbers

which results in numerical instabilities.

2

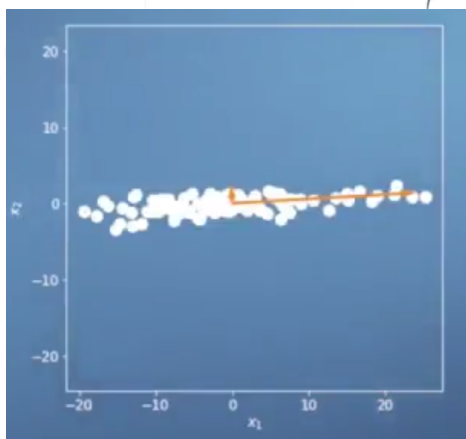
② ~~and add~~

② second step which is normally recommended after subtracting the mean is to divide every dimension of the centered data by the corresponding standard deviation.

This makes the data unit free and guarantees that the variance of the data in every dim $\rightarrow 1$.

But it leaves the correlations intact.

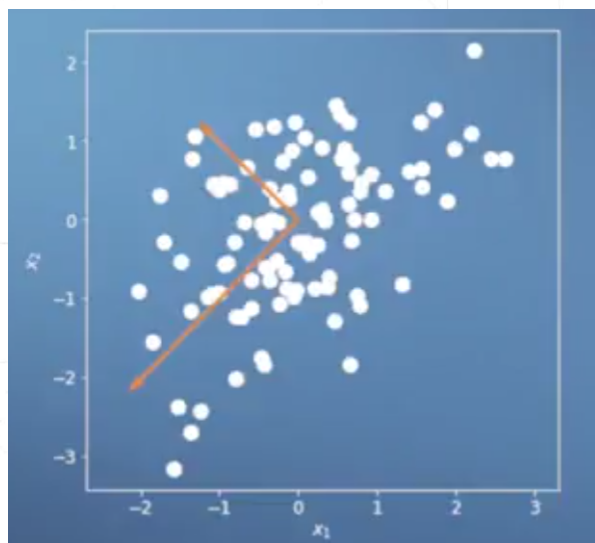
Let's have a look at an example (see pic)



Clearly the data spreads much more in x_1 than other dim., and the Best projection of PCA is clear

However there is a problem, with this data set. 3
The two dims of data set are both different,
But one is measured in ~~cm~~ cm, and other in meter.
These measured in cm, naturally variance much
more than the other one.

When we divide each Dim of dataset
by the corresponding std dev, we get
rid of the units and make sure that
the variance in each Dim is 1. (scale)

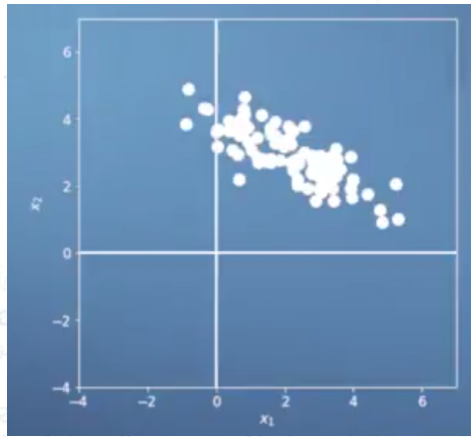


When we look at the principal subspace of this
normalized dataset, we can now see
that there is a clearly quite a strong
Correlation between these two dimensions.

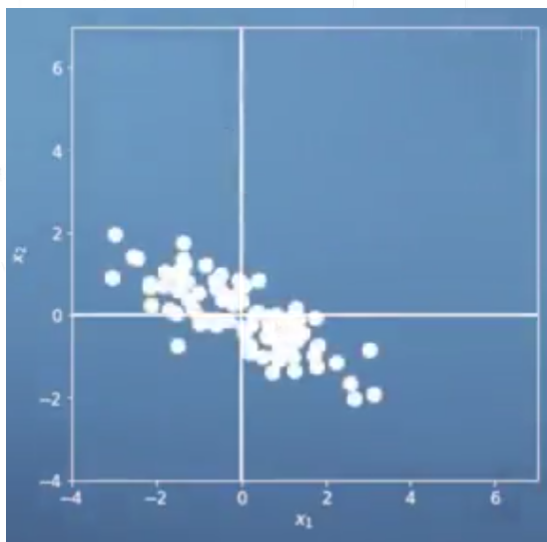
And the principal axis has changed. But
But now let's go through PCA, step by step,
and we'll have a running example.

(see pic)

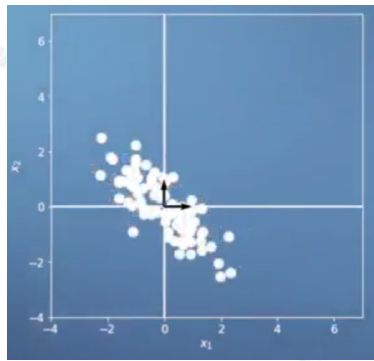
we are given a two dim data set and we
want to use PCA to project it onto a
1 Dim subspace (see pic)



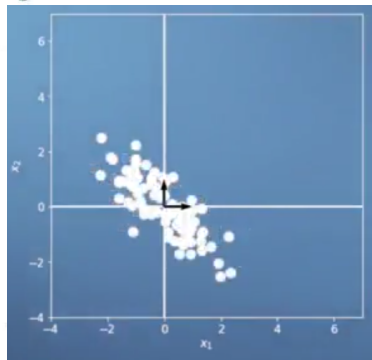
The first thing that we do \rightarrow to subtract the
means (see pic)



The data is now ~~set~~ centered.
Next we divide by the std dev. (see pic)

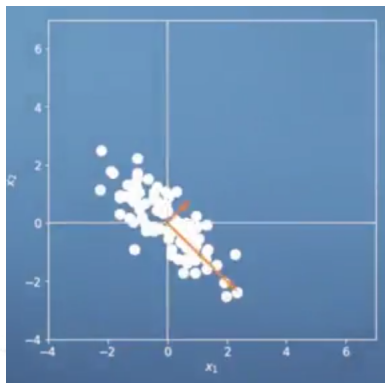


Now the data is unit free, and has ~~variance~~ 1,
and has 1 along each axis, which is
indicated by these two arrows (see pic)



But keep in mind that the Correlations are
still intact.

Third, we compute the data Covariance
matrix, and its eigenvalues and
Corresponding eigenvectors. (see pic)



The eigenvectors are scaled by the magnitude of the corresponding eigenvalues in the PR. The larger vector spans the principal subspace. Let's call it U_1 .

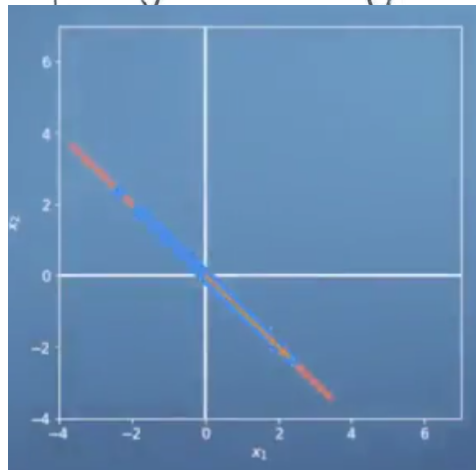
and last step, we can project any datapoint, X_i onto the principal subspace.

To get this right we need to normalize X_i first, using the mean and std dev of data set, that we use to compute the data Covariance matrix.

"So we going to have a new X_i^* , and the new X_i^* is going to be old X_i minus the mean of data set, divided by std dev and we do this for every dim in X_i^* "

$$X_*^{(d)} \leftarrow \frac{X_*^{(d)} - \mu^{(d)}}{\sigma^{(d)}}$$

Now we can get the projection of X_* (see pic)



"This is \tilde{X}_* or the projection of X_* onto the principal

subspace U as B times B transpose X_* ,

where B is matrix that contains the eigenvectors that belongs to the largest eigen values

as columns, and B^T times X_* is the coordinates of the projection w.r.t the basis of the principal subspace"

$$\tilde{X}_* = \Pi_U(X_*) = B(B^T X_*)$$

In this session, we through the steps of PCA.

First, we subtract the mean from data and centered it 0, to avoid numerical problems

Second, we divide by the std dev, to make the data unit free.

Third, we compute the eigenvalue / vector of basis data Co-variance matrix

Finally, we can project any data point onto the principal subspace that's spanned by the eigenvectors that belong to the largest eigenvalues