Module 4

Summary

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

Before we do this lets 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 a 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 and addth

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 is 1.

But it leaves the correlations in tact.

Lets have a look at an example (see pic)

Clearly this data spreads much more in 1 Dim than other dim., and the Best projection of PCA is clear

However there is a problem, with the data ~~stet~~ set.
The two dims of data set are both distances,
But one is measured in ~~cm~~ cm, and other in meters.
Those measured in cm, naturally vary much
more than the other one.
When we divide each Dim of dataset
by the corresponding std dev, we get
rid of the unit and make sure that
the variance in each Dim is 1. (see pic)

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

And the principal axis has changed. ~~but~~
But now lets go through PCA, step by step,
and we'll have a running example.
(see pic)
We are giving 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 - to subtract the
mean (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 PC.
The larger vector spans the principal subspace, lets call it U,

and last step, we can project any datapoint, X_*^* onto the principal subspace

To get this right we need to normalized X star, using the mean and std dev of data set, that we are to compute the data covariance matrix.

"So we going to have a new X_*, and the new X_* is going to be old X_* minus the mean of data set, divided by std dev and we do this for every dim $^{(d)}$ in X_*"

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

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

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

In this session, we through the steps of PCA.

First, we subtract the mean from data and centred 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 is spanned by the eigenvectors that belong to the largest eigenvalues