Module 3 & 4

Motivation:

In this session we going to introduce Principal Component analysis (PCA), an algorithm for linear dimensionality reduction.

PCA has been around for about 100 years, and is still one of the most commonly used techniques for data compression and visualization.

High Dim data for eg, images often has the property that it lies on a low Dim subspace, and that many Dims are highly correlated.

Here is an illustration in 2 Dims (see pic)

Although the data doesn't quite lie in a straight line, the data does not vary much in one direction.

so that we can express it as if it was on a line with nearly no loss (see pc)

A key idea behind PCA is to use orthogonal projections to find lower dim representation of data that retain as much information as possible

Similar to the example we just looked at.

In the following sessions we will derive PCA as an algorithms that minimises average reconstruction error by orthogonal projections