☰ **Item Navigation**

PCA algorithm

PCA is commonly used to reduce the dimension of your data. Intuitively the model collapses the data across principal components. You can think of the first principal component (in a 2D dataset) as the line where there is the most amount of variance. You can then collapse the data points on that line. Hence you went from 2D to 1D. You can generalize this intuition to several dimensions.
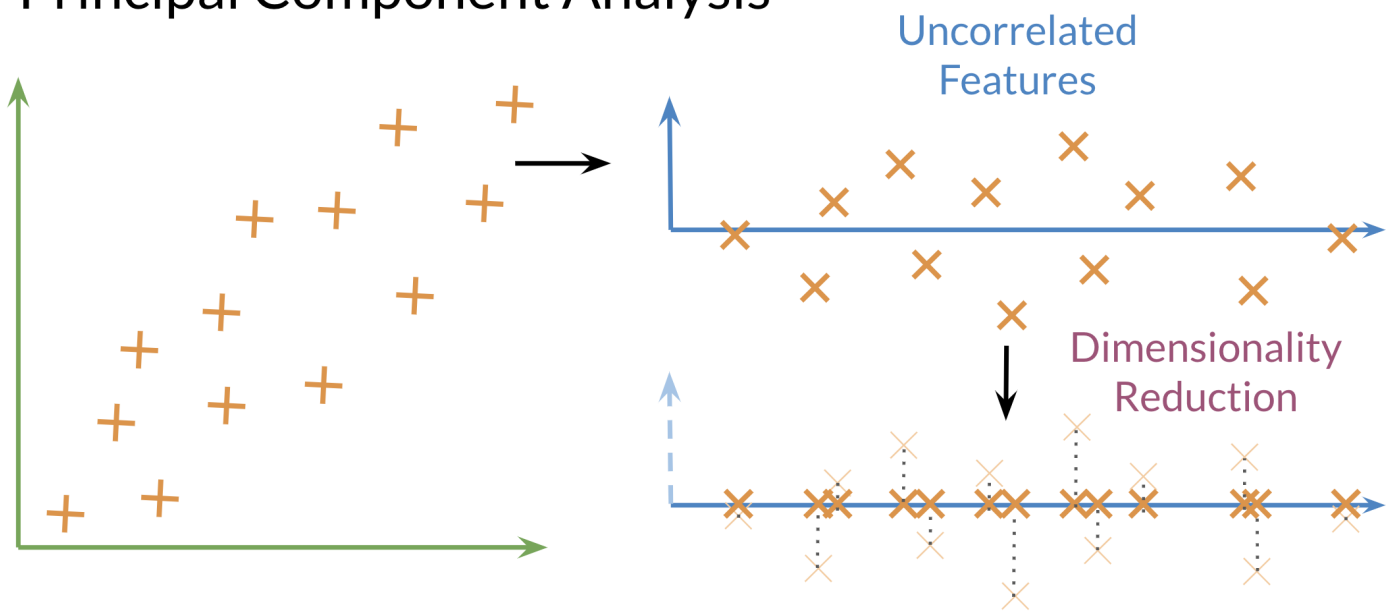

Eigenvector: the resulting vectors, also known as the uncorrelated features of your data

Eigenvalue: the amount of information retained by each new feature. You can think of it as the variance in the eigenvector.

Also each **eigenvalue** has a corresponding eigenvector. The eigenvalue tells you how much variance there is in the eigenvector. Here are the steps required to compute PCA:
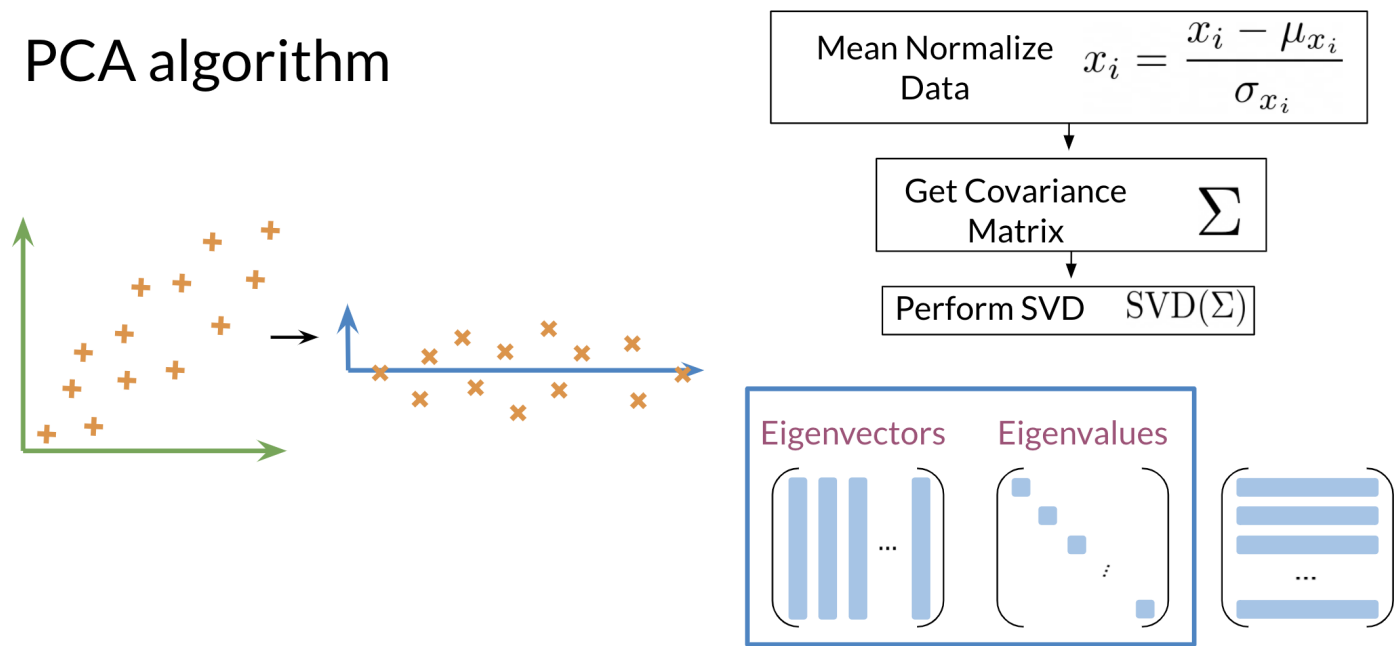
Steps to Compute PCA:

- Mean normalize your data

- Compute the covariance matrix

- Compute SVD on your covariance matrix. This returns $[USV] = svd(\Sigma)$. The three matrices U, S, V are drawn above. U is labelled with eigenvectors, and S is labelled with eigenvalues.

- You can then use the first n columns of vector U, to get your new data by multiplying $XU[:, 0 : n]$.