PCA

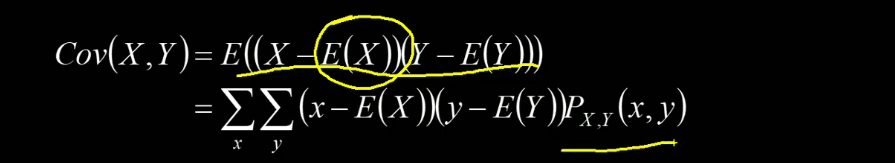
Dimension reduction method

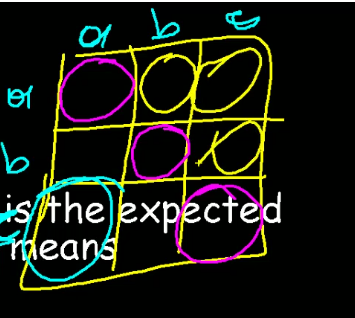
In machine learning having too many columns

Visualization of high dimensional data

Curse of dimensionality – need exponential number of samples as dimensionality increases

Take a many dim. dataset and transform it to a lesser dimension without loosing too much information contained in the many dimensional dataset

Covariance - how two variablesmove/change together



Covariance matrix – variance on diagonal, covariances in other cells, symmetrical matrix.

Non-linear relationships between the variable can get lost when performing PCA