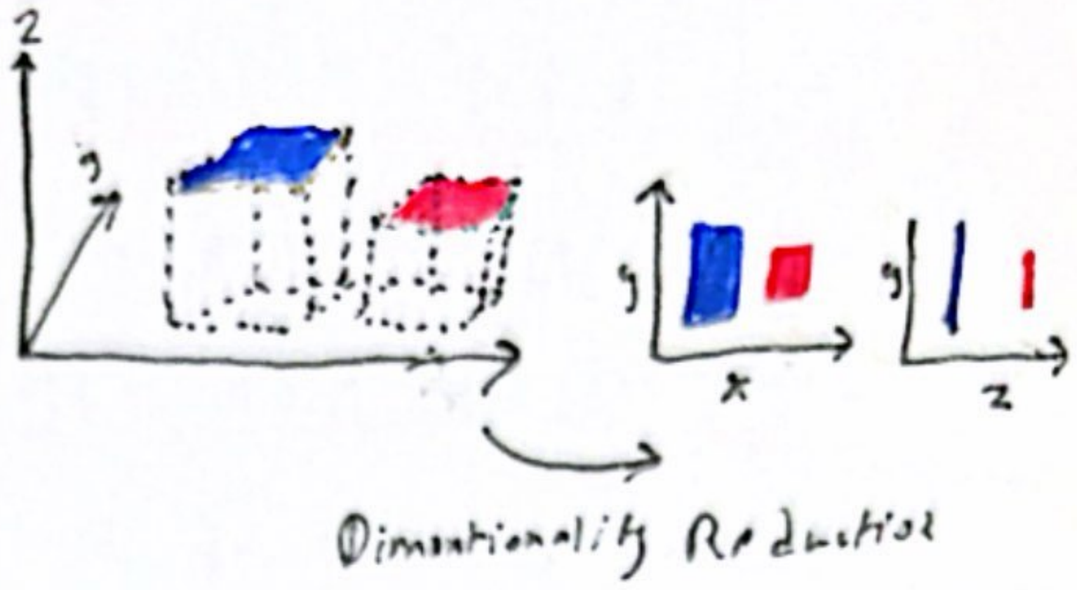
(see Autoencoders too)

# Dimensionality Reduction

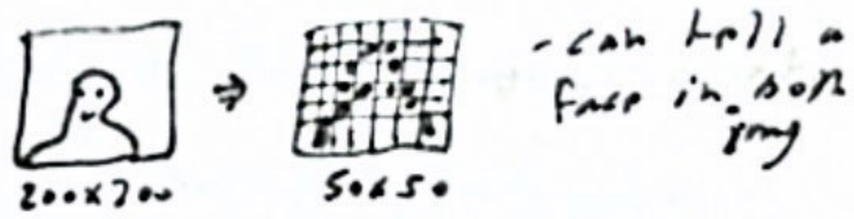
- the idea of this is to reduce the number of features or Dimentions of your data set while keeping as much info as possible




- Usally this group of algos do this by finding correlations between existing features and removing potentially redundant features without losing much information

Ex do you really need a picture in high resolution to know this image has a face?
or can you reduce num of pixels (features/Dim)

- this can also be used as a data preprocessing step for supervised learning algo to make algo more efficient




- can tell a face in 50x50 img

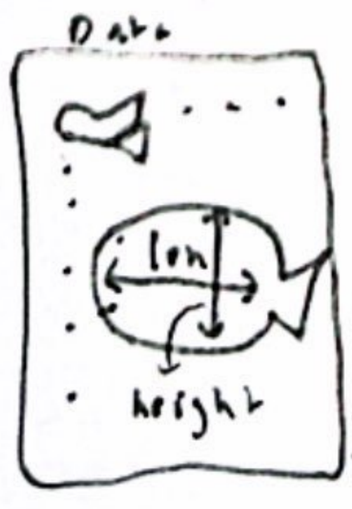
## Principal Component Analysis (PCA)

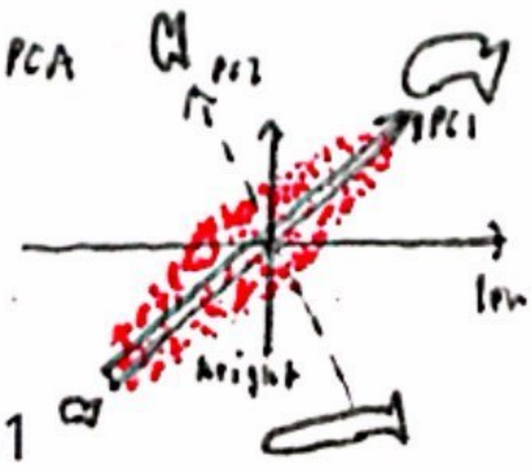
- is one algorithm for Dimentionality reduction lets say we are trying to predict fish on features like len, height, color and number of teeth. when looking at correlations of different features we find height and length are strongly correlated and including both wont help the alg much and might hurt it by introducing noise so we can simply include a shape feature that is a combination of the two this is popular in large data sets it allows us to reduce dimentions while keeping info

* PCA does this by finding the direction in which most variance in data set is retained

- in this ex direction of most variance is a diagonal called the PC #1 and is new shape feature

- the PC #2 is normal to the first explains small fraction of variance and can be ignored




- Do this for all Features Rank and Remove the lowest variance