Matrix Factorization Methods

Principal Component Analysis

↳ Collaborative filtering works well for very large scale data.

↳ If data is sparse / for small data its not good to do Collaborative filtering ~~because for small data~~ getting similarity matrix as ~~matrix~~, it is difficult to do get Similarity matrix for Collaborative filter (as the data is small)

↳ another negative for Collaborative is it takes lot of Computation power.

↳ In Item-based Collaborative filtering, it reduces the Complexity of that above problem; So we can Compute Similarity matrix easily (but not tough)

{ Items are normally less than people }

{ more computation power }

↳ But the main problem is, it is sensitive to noisy data & sparse data.

↳ In order to avoid these lets explore some other methods to make recommendations. → model-based methods

↳ Instead of trying to find items/users that are similar to each other, we will use data science & ML techniques to extract predictions from our ratings data.

Matrix Factorization

↳ In ML algorithms, we will convert items into matrices

(~~Comedy~~ action, romantic, ...)

↳ generally we will describe items in amounts of each future

{ i.e Bob is defined as 80% action & 20% Comedy }

movies

users	Indian jones	Star wars	XYZ	BB1	BB2
Bob	4	5	?	?	?
Ted	?	?	?	?	1
Ann	?	5	5	5	?

↳ Data is sparse

↳ Here, we want to fill the unknown cells with predictions

↳ These Missing Values are predicted using ~~Matrix Factorization~~ ML Techniques like PCA ~~to predict the predictions~~

↳ ~~One such technique is called~~

PCA (principal component analysis)

↳ PCA is usually described as dimensionality reduction problem.

↳ PCA will Convert ~~do like~~, data that exists in different directions into smaller set of dimensions

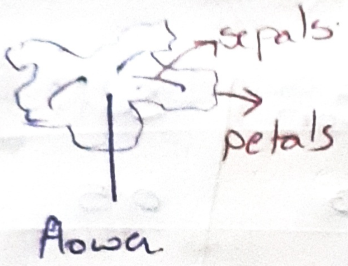
↳ Extracts features from Data.

↳ Eigen vector are principal components & thats why its called as PCA

↳ ~~These Eigen vectors are ...~~

↳ Here, we will find principal components that describes our data

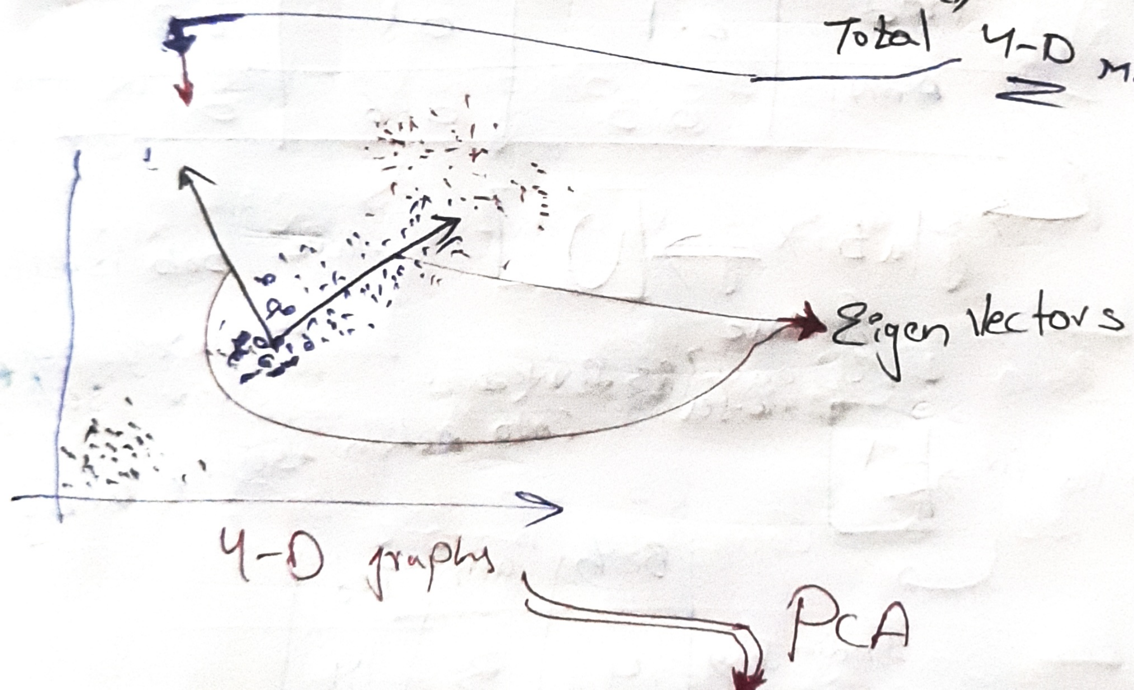
Ex: flower.
(Iris dataset)



Petals → length & width
(l_1) (w_1)

Sepals → " "
(l_2) (w_2)

⇒ Total 4-D Matrix



PCA :-
Extracts features from Data.