Embedding Types and Uses

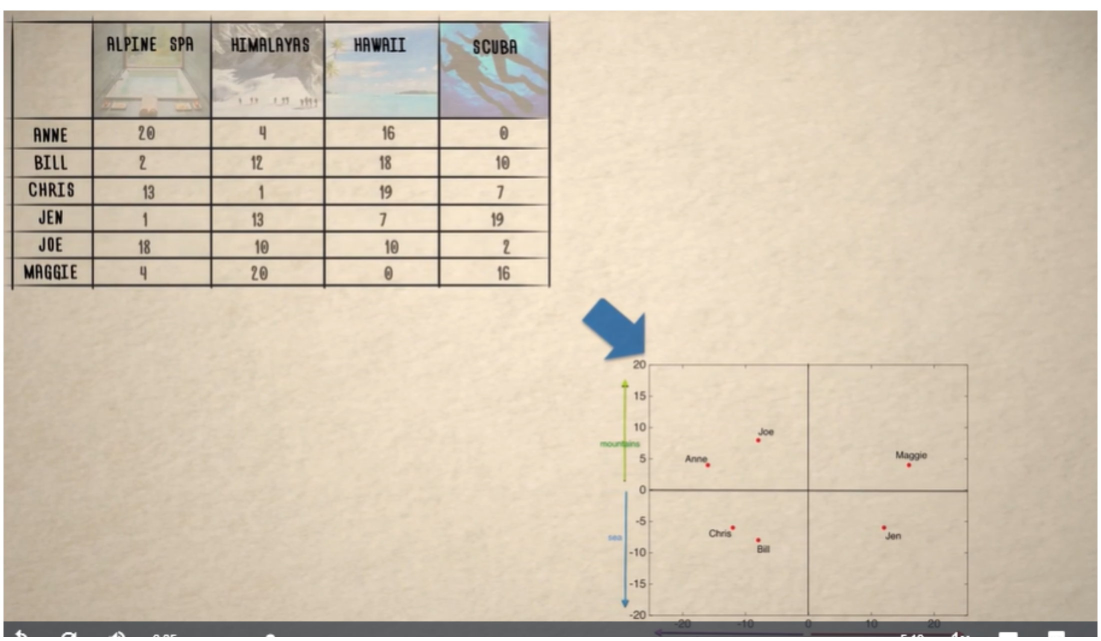
In the previous sections, we saw PCA, embedding, and Spectral / Modularity clustering. All these methods find new feature vectors for each data point in the graph, which we also call embeddings.

The idea of embedding is very useful as it helps us to create new features based on the unstructured data also and thus helps in finding hidden patterns in the data.
Let’s now look in brief at different types of embeddings and their uses.

With the help of eigenvectors we saw, different types of embedding can look very different from each other and can carry different patterns.

First, we used eigenvectors to create embedding and did dimensionality reduction using PCA, and we then saw we can use these embedding to visualize the results and find meaningful groups.

On the left side, we have our data from the holiday destination and we applied PCA and reduced it to 2 dimensions, and then visualize the results.

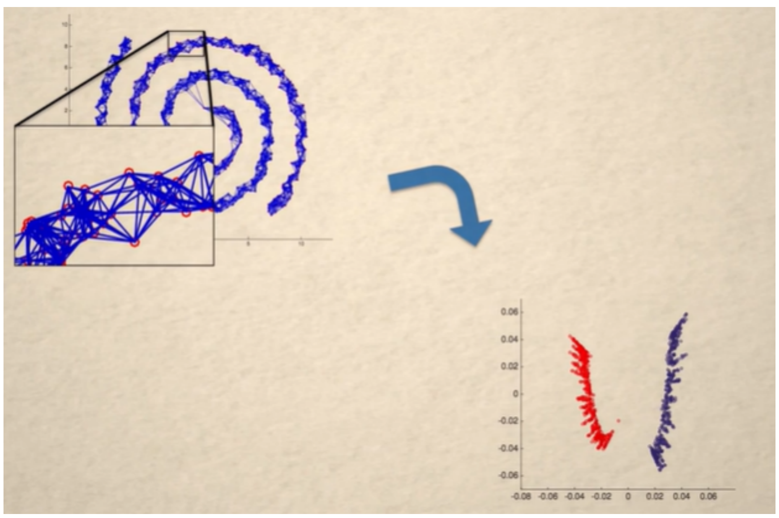


	ALPINE SPA	HIMALAYAS	HAWAII	SCUBA
ANNE	20	4	16	0
BILL	2	12	18	10
CHRIS	13	1	19	7
JEN	1	13	7	19
JOE	18	10	10	2
MAGGIE	4	20	0	16

As we have seen in PCA, these embedding can give important components or themes, this points to a difference between a spectral embedding and PCA.

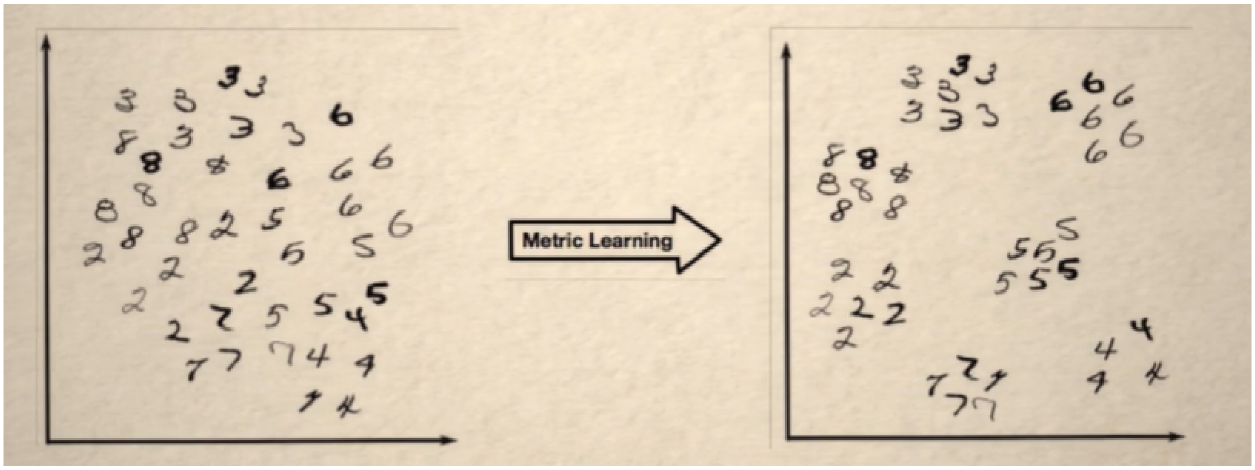
In PCA, each datapoint can be written as the weighted sum of the components; such embeddings are called linear embeddings(as we can represent them as a weighted sum of components)

In Spectral Clustering, we first construct the graph and then find eigenvectors and even if we start out with feature vectors our representation is not linear, such embeddings are called non-linear embeddings.

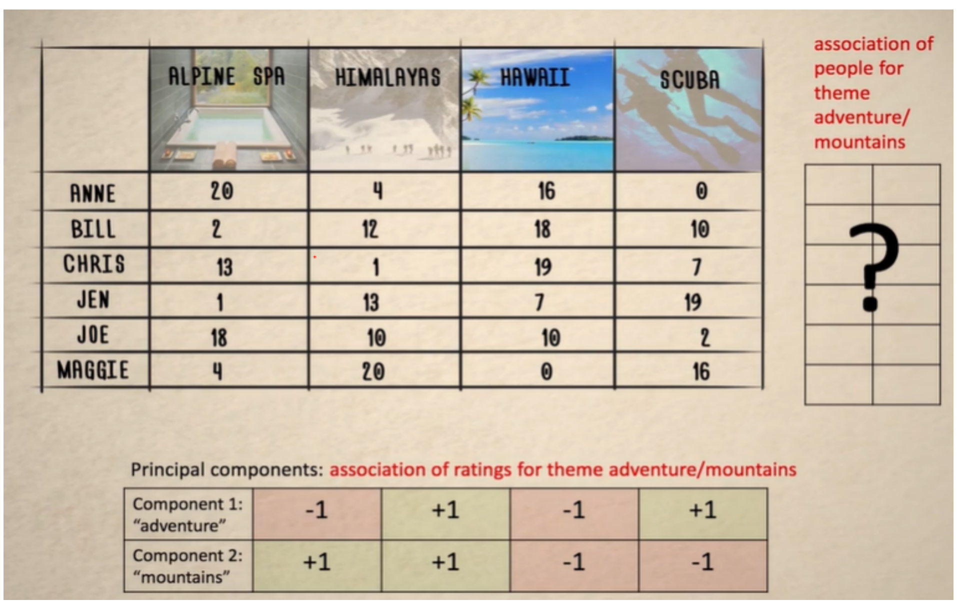


Now, we say that different types of embedding are made based on different goals and different criteria. For example, in spectral clustering, we only knew the pairwise similarity between the nodes, like friendship in the social network, and related articles for given articles in a collection. Our goal is to find a vector representation that can maintain the local relationships and put related points close to each other.
This goal of capturing local relationships and putting them together is used by many other methods.

For example - metric learning, in this we have some side information to guide the embeddings, say we have images of handwritten digits and we are aware of some correct digits, for some of them. Now we want an embedding that can separate all the digits, we can guide the embedding so that the example pairs should be close because they are the same digit. Thus securing the local relationship can be achieved by metric learning, which takes the side information and guides the embedding.



Now, let's look at another important concept that is closely related to PCA. It is called **Singular Value Decomposition (SVD)**. This can also give us the principal components.



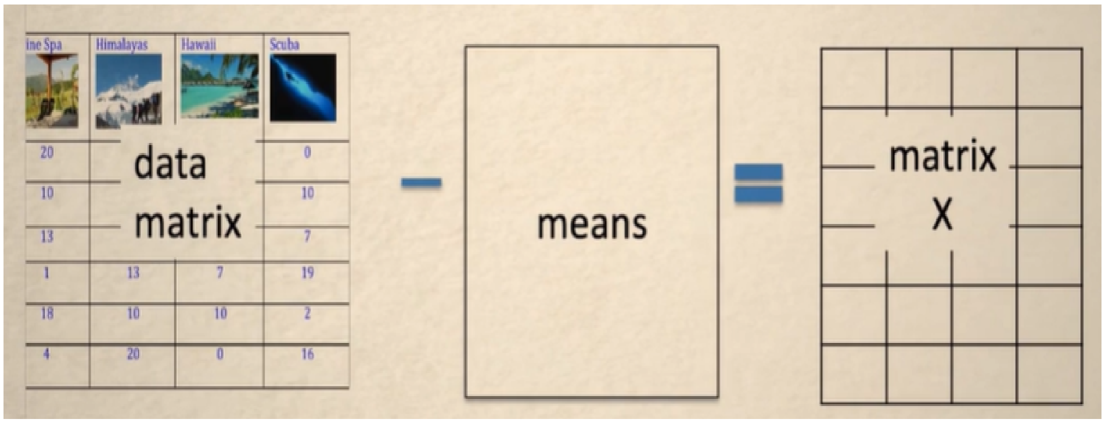
	ALPINE SPA	HIMALAYAS	HAWAII	SCUBA
ANNE	20	4	16	0
BILL	2	12	18	10
CHRIS	13	1	19	7
JEN	1	13	7	19
JOE	18	10	10	2
MAGGIE	4	20	0	16

Principal components: association of ratings for theme adventure/mountains

Component 1: "adventure"	-1	+1	-1	+1
Component 2: "mountains"	+1	+1	-1	-1

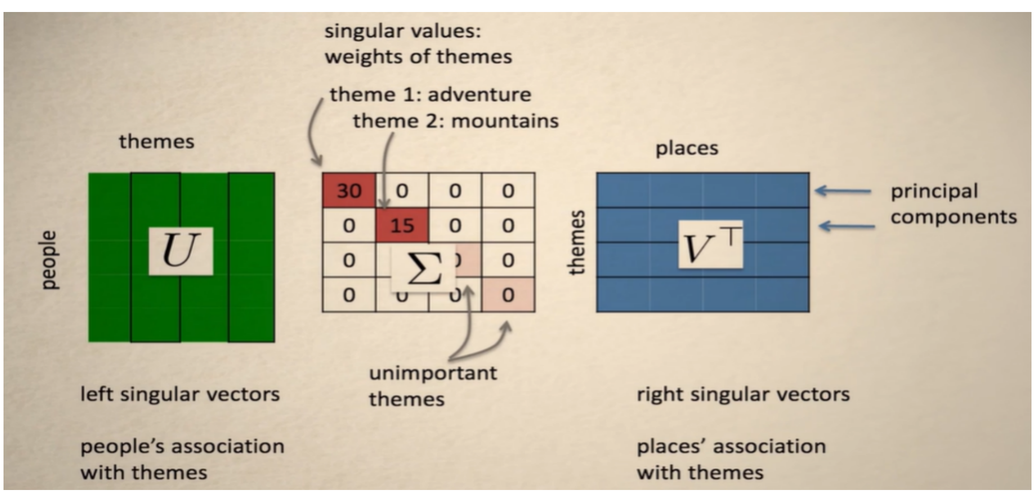
Recall the PCA example: People rate the holiday places and each row in the data matrix corresponds to 1 person and each column corresponds to the destination and each principal component is a pattern like adventure and mountain and associated places with those examples, but in SVD, we can get those associations for both people and the places.

Let's look at its working:
We take the data matrix and subtract the mean, this gives us the X.



In SVD, write the data matrix as the product of 3 matrices, which are usually called U, Sigma, and V^T. The sigma matrix is all 0 except for the diagonal. The squares of the diagonal entries in sigma are values we get in PCA. The role of V is the same as the eigenvectors we get in PCA.

This is also called the **Right singular vector** as it is in the rightmost.



These 3 Matrices describe our data via theme, for each theme we have a singular value, a right singular matrix, and a left singular matrix. The singular values tell the importance of the theme as they have eigenvalues on the diagonals. The right singular vector shows the association of each place with a theme. Recall, that scuba diving was associated with adventure, but not mountains. The corresponding left singular vector tells the preference of each person for a particular pattern, so we can simultaneously understand the liking of people and the characteristics of the places. The entries in the left vectors are multiplied by the respective singular values and are exactly the coordinates we use to visualize the data.