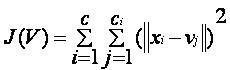
## Eigenface

**Eigenfaces** is the name given to a set of [eigenvectors](https://en.wikipedia.org/wiki/Eigenvector) when they are used in the [computer vision](https://en.wikipedia.org/wiki/Computer_vision) problem of human [face recognition](https://en.wikipedia.org/wiki/Facial_recognition_system).the presence of some objects (eyes, nose, mouth) in any face as well as relative distances between these objects. These characteristic features are called *eigenfaces*in the facial recognition domain (or *principal components*generally). They can be extracted out of original image data by means of a mathematical tool called *Principal Component Analysis* (PCA).  The eigenvectors are derived from the [covariance matrix](https://en.wikipedia.org/wiki/Covariance_matrix) of the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) over the high-[dimensional](https://en.wikipedia.org/wiki/Dimension) [vector space](https://en.wikipedia.org/wiki/Vector_space) of face images.

**K-means**

k-means is  one of  the simplestunsupervised  learning  algorithms  that  solve  the well  known clustering problem.  The  main  idea  is to define k centers, one for each cluster. These centers  should  be placed in a cunning  way  because of  different  location  causes different  result. this  algorithm  aims at  minimizing  an objective function know as squared error function given by:

[](https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm/kmeans.JPG?attredirects=0)

where,  
                           *‘||xi- vj||’* is the Euclidean distance between *xi* and *vj.*

*‘ci’* is the number of data points in *ith* cluster.

*‘c’* is the number of cluster centers.

**Advantages**

1) Fast, robust and easier to understand.

2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t  is # iterations. Normally, k, t, d << n.

3) Gives best result when data set are distinct or well separated from each other.

**Disadvantages**

1) The learning algorithm requires apriori specification of the number of  cluster centers.

2) The use of  Exclusive Assignment - If  there are two highly overlapping data then k-means will not be able to resolve       that there are two clusters.

3) The learning algorithm is not invariant to non-linear transformationsi.e.with different representation of data we get

    different results (data represented in form of cartesian co-ordinates and polar co-ordinates will give different results).

4) Euclidean distance measures can unequally weight underlying factors.

5) The learning algorithm provides the local optima of the squared error function.

6) Randomly choosing of the cluster center cannot lead us to the fruitful result. Pl. refer [Fig](https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm/k-means_initial_cluster_selection).

7) Applicable only when mean is defined i.e. fails for categorical data.

8) Unable to handle noisy data and outliers*.*