Dimensionality Reduction:

Dimensionality reduction is simply, the process of reducing the dimension of your feature set. Your feature set could be a dataset with a hundred columns (i.e. features) or it could be an array of points that make up a large sphere in the three-dimensional space. Dimensionality reduction is bringing the number of columns down to say, twenty or converting the sphere to a circle in the two-dimensional space.

Purpose of Dimensionality Reduction:

1. More than 3 dimensions of data, can’t visualize .so for visualization purpose reduce dimension of data.
2. Less dimensions mean less computing. Less data means that algorithms train faster.
3. Removes redundant features and noise.
4. Less misleading data means model accuracy improves.
5. Less data means less storage space required.
6. Less dimensions allow usage of algorithms unfit for a large number of dimensions.

Methods of Dimensionality reduction:

**1.Linear Dimensionality reduction methods**

**PCA (Principal component analysis):** Popularly used for dimensionality reduction in continuous data, PCA rotates and projects data along the direction of increasing variance. The features with the maximum variance are the principal components.

**LDA (Linear Discriminant Analysis):** projects data in a way that the class separability is maximised. Examples from same class are put closely together by the projection. Examples from different classes are placed far apart by the projection.

**2. Non-linear Dimensionality Reduction Methods**

**Isometric Feature Mapping (Isomap) :** Projects data to a lower dimension while preserving the geodesic distance (rather than Euclidean distance as in MDS). Geodesic distance is the shortest distance between two points on a curve.

**Locally Linear Embedding (LLE):** Recovers global non-linear structure from linear fits. Each local patch of the manifold can be written as a linear, weighted sum of its neighbours given enough data.

**t-distributed Stochastic Neighbour Embedding (t-SNE):** Computes the probability that pairs of data points in the high-dimensional space are related and then chooses a low-dimensional embedding which produce a similar distribution.