Dimensionality :

No. of columns , features, input variables, dimensions in input space.

* Beyond a point , a greater number of features starts hurting more than helping

**Formal Techniques for reducing dimensions of feature seta are:**

1. PCA
2. Factor Analysis
3. Linear Discriminant Analysis

**Curse of Dimensionality:**

* Problems arising only in high dimensional space

The Problem :

* As dimensionality increases, the volume of feature space increases exponentially.
* Space becomes sparse
  + Much lesser data density.
* Sparse spaces don’t give statistically significant results.
* Very difficult to find groups of data points(by similarity) in sparse spaces
  + All data points are far apart from each other
  + Classifiers like knn would fail
* Number of samples would have to be increased exponentially to get useful results.

**PCA Algorithms:**

* Fit Lines to given data
  + N lines for n dimensions
* Each subsequent line is orthogonal in direction to all previous ones.
* Each of these lines: Principal Components
* Principal components have been arranged in order of importance by using this techniques.
* Pick the first p of these n lines
* Dimensions reduced from n to p
  + Some loss of information.
  + How is the data spread, i.e., variability.