11_Dimensionality Reduction

```
Concept of subspace
Similarity measures
Subspace methods
Fisher's criterion
```

Principle Component Analysis (PCA)

Two main equivalent ideas: Maximizing variance & Minimizing approximation error.
choose largest variance orientation

PCA Algorithm

Input:

Sampling dataset : $D = \{\vec{x}_1, \vec{x}_1, \dots, \vec{x}_m\}$

dimensions in lower dimensional space : d'

Procedure:

1. Centralization all sampling datas: $\vec{x}_i = \vec{x}_i - \frac{1}{m}\sum_{i=1}^{m} \vec{x}_i$
2. Compute covariance (diversity of sampling data distribution) matrix: $\vec{X}\vec{X}^T$
3. Eigen decomposition covariance matrix
4. Take maximal d' eigenvalues as corresponding eigenvector $\vec{\omega}_1, \vec{\omega}_2, \dots, \vec{\omega}_{d'}$.

Output:

$W = \{\vec{\omega}_1, \vec{\omega}_2, \dots, \vec{\omega}_{d'}\}$

Subspace

Reference: https://www.youtube.com/watch?v=_UVHneBUBW0