Unsupervised Learning: Dimensionality Reduction

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Dimensionality Reduction: An Overview

- **Why Dimensionality Reduction ?**
- Consider a data set that contains image of letter A which has been scaled and rotated. each image of A is a 32×32 pixel image so it would be a 1024 dimensional data.

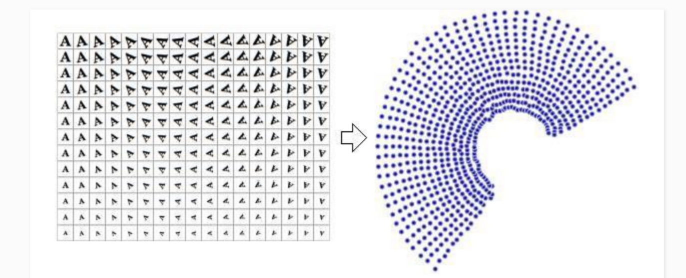

Figure: Dimensionality Reduction, Source

Dimensionality Reduction: An Overview

- However, in the preceding picture, it looks like that the actual dimension of the data is two because only two variables namely rotation and scale used to generate the data set.
- A successful dimensionality reduction progress would extract the variable information (rotation and scale) and discard the correlated information (letter A)
- Unlike clustering, which we searched for the discerte latent variables, such as number of clusters and the behavior of them, in dimensionality reduction we would like to find the **continuous latent variables** . these important features are also called **degrees of freedom** .
- In general, we would like to discover another data space which has much lower dimensionality than the original data space and these data are closed to it.

An Overview of PCA

- One of the most widely used techniques in order to do the dimensionality reduction is **Principal Comppponent Analysis** or **PCA** .
- Principal Component Analysis is usually defined in two ways, although these two definitions are eqauivalent:
  - **Maximum Variance Formulation**
  - **Minimum Error Formulation**

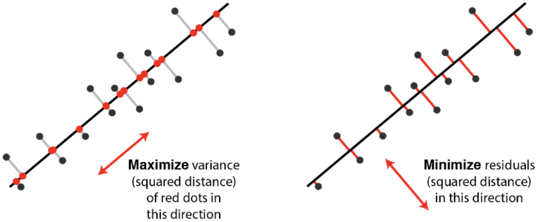

Figure: Illustration of PCA, Source

PCA in Detail: Maximum Variance Formulation

- Consider we have a D dimensional data set $\{\mathbf{x}_n\}$, which we would like to reduce the number of dimensions to the $M(M < D)$. without loss of generality we assume that the $M = 1$. we then generalize the case of M-dimensional by induction.
- We would like to compute the variance of projected data and then maximize it, if we assume that the direction of real line in D-dimensional space is shown by a unit vector called $\mathbf{u_1}$, then the emprical covariance of projected 1-Dimensoional data would be:

$$
\begin{gathered}
Cov = \frac{1}{N-1} \sum_{n=1}^{N} \left(\mathbf{u_1}^T \mathbf{x}_n - \mathbf{u_1}^T \bar{\mathbf{x}} \right)^2 = \mathbf{u_1}^T \mathbf{S} \mathbf{u_1} \\
\text{where} \quad \bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \quad , \quad \mathbf{S} = \frac{1}{N-1} \sum_{n=1}^{N} (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T
\end{gathered}
\tag{1}
$$

PCA in Detail: Maximum Variance Formulation

- In order to maximize the variance we would have a constrained optimization problem like this:

$$\max_{\mathbf{u_1}} \quad \mathbf{u_1}^T \mathbf{S} \mathbf{u_1}$$
$$\text{subject to} \quad 1 - \mathbf{u_1}^T \mathbf{u_1} = 0 \tag{2}$$

- If we use the Lagrangian multiplier we have an unconstrained optimization:

$$\max_{\mathbf{u_1}} \quad \mathbf{u_1}^T \mathbf{S} \mathbf{u_1} + \lambda_1 (1 - \mathbf{u_1}^T \mathbf{u_1}) \tag{3}$$

- Taking deravitive of preceding term respect to $\mathbf{u_1}$ and set it to zero would give us:

$$\mathbf{S} \mathbf{u_1} = \lambda_1 \mathbf{u_1} \tag{4}$$

- It is obvious that the $\mathbf{u_1}$ must be the **eigenvector** of $\mathbf{S}$ and the λ_1 is the corresponding **eigenvalue** , this eigenvector is known as **first principal component** . if we left-multiply both side of above equation by $\mathbf{u_1}^T$, the maximum variance is given by:

$$\mathbf{u_1}^T \mathbf{S} \mathbf{u_1} = \lambda_1 \tag{5}$$

PCA in Detail: Singular Value Decomposition

- As a result, we must take the $\mathbf{u_1}$ which has the **largest** λ_1 value corresponding to it.
- We can consider the general M-dimensional case by obtaining $\mathbf{u_1}, \mathbf{u_2}, \dots, \mathbf{u_m}$ corresponding to the first M large eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$. this can easily shown by induction
- To summerize, PCA is dealing with the covariance matrix of the data set and searching for M largest eigenvalues corresponding to the M eigenvectors that form the basis of the final subspace.

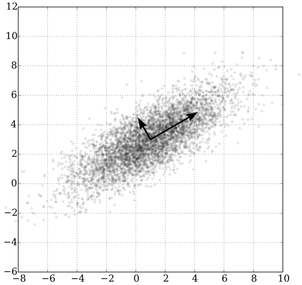

Figure: Illustration of PCA, Source

PCA in Detail: Singular Value Decomposition

- There are several techniques to decompose the eigenvectors, but the most useful solution is **Singular Value Decomposition** .

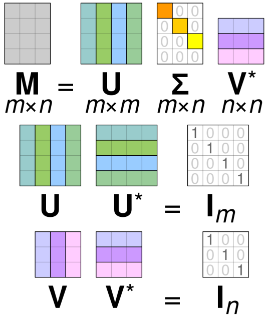

Figure: Illustration of SVD, Source

Thank You!

Any Question?