PCA

Two ways to derive PCA

① Maximize variance after projection

② Minimize Reconstruction error

We show ① and ② are the same

① $X \in \mathbb{R}^{n \times p}$ data matrix, each row a data point

Assume data points are centered, i.e.

$\bar{x}^T = \frac{\mathbb{1}^T X}{n} = \vec{0}$, $\mathbb{1}$: a vector of ones

Projection of X onto $\vec{v} \in R^p$: $X\vec{v} \in \mathbb{R}^n$

Sample mean of $X\vec{v}$: $\frac{\mathbb{1}^T X \vec{v}}{n} = \bar{x}^T \vec{v} = \vec{0}^T \vec{v} = 0$

Sample variance of $X\vec{v}$: $\frac{1}{n-1}(X\vec{v})^T(X\vec{v}) = \frac{1}{n-1}\vec{v}^T X^T X \vec{v}$

Find $\vec{v}$ that maximizes the variance.

$$\max_{\vec{v}} \ \vec{v}^T X^T X \vec{v} \quad \text{s.t.} \quad \vec{v}^T \vec{v} = 1$$

Solution: solve eigenvalue problem $X^T X \vec{v} = \lambda \vec{v}$ ...... ①

Among all $(\lambda, \vec{v})$ that satisfies ①, choose the one whose λ is the largest since $\vec{v}^T X^T X \vec{v} = \lambda \vec{v}^T \vec{v} = \lambda$

(2)

$$X = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix} \begin{matrix} \to \text{first data point} \\ \\ \to n^{th} \text{ data point} \end{matrix}$$

$(X\vec{v})_i = \vec{x}_i^T \vec{v}$ is the projection of $\vec{x}_i$ onto $\vec{v}$

$(X\vec{v})_i \vec{v}$ is the reconstruction of $\vec{x}_i$ using $\vec{v}$

Reconstruction error: $\sum_{i=1}^{n} \| \vec{x}_i - (X\vec{v})_i \vec{v} \|^2$

$$= \sum_{i=1}^{n} \left(\vec{x}_i^T \vec{x}_i - 2(X\vec{v})_i \vec{x}_i^T \vec{v} + ((X\vec{v})_i)^2 \vec{v}^T \vec{v} \right).$$

We require $\vec{v}^T \vec{v} = 1$

$$= \underbrace{\sum_{i=1}^{n} \vec{x}_i^T \vec{x}_i}_{\text{constant}} - \sum_{i=1}^{n} ((X\vec{v})_i)^2$$

So, minimizing the Reconstruction error is the same

as $\min_{\vec{v}} \; - \sum_{i=1}^{n} ((X\vec{v})_i)^2 = - \| X\vec{v} \|^2$

$$= - \vec{v}^T X^T X \vec{v}$$

s.t. $\vec{v}^T \vec{v} = 1$

which is equivalent to

$$\max \; \vec{v}^T X^T X \vec{v} \quad \text{s.t. } \vec{v}^T \vec{v} = 1$$

Laplacian Eigenmap and Spectral Clustering

n data points $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$

$A \in \mathbb{R}^{n \times n}$, $A^T = A$ an affinity matrix

A_{ij} : similarity between $\vec{x}_i$ and $\vec{x}_j$

D : a diagonal matrix such that $D_{ii} = \sum_{j} A_{ij}$

Laplacian $L = D - A$

Laplacian Eigenmap

$$\min \sum_{i,j} A_{ij}(f_i - f_j)^2 = f^T L f$$

one dimensional representation of data points

$$\text{s.t. } f^T D f = 1$$

Spectral Clustering (Relaxed Normalized Cut)

$$\min \frac{f^T L f}{f^T D f} \quad \text{s.t. } f^T D \mathbb{1} = \vec{0}$$

Obtain cluster assignments by thresholding f

$\Longleftarrow$ Require $f^T D f = 1$ (remove arbitrary scaling)

Require $f^T D \mathbb{1} = \vec{0}$ ($f = \mathbb{1}$ is a trivial solution) $\Longrightarrow$

Symmetrically normalized Laplacian:

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \quad (\text{Assume } D^{-1} \text{ exists})$$

$$\min y^T \tilde{L} y \quad \text{s.t. } y^T y = 1$$

$$\min \frac{y^T \tilde{L} y}{y^T y} \quad \cancel{\text{s.t. } y^T \mathbb{1} = \vec{0}}$$

$$y^T D^{\frac{1}{2}} \mathbb{1} = 0.$$

~~Let~~

Variable transformation:

$$D^{-\frac{1}{2}} y = f \iff y = D^{\frac{1}{2}} f$$

well-defined if D^{-1} exists.

Then

$$y^T \tilde{L} y = f^T D^{\frac{1}{2}} D^{-\frac{1}{2}} L D^{-\frac{1}{2}} D^{\frac{1}{2}} f = f^T L f$$

$$y^T y = 1 \iff f^T D^{\frac{1}{2}} D^{\frac{1}{2}} f = f^T D f = 1$$

$$y^T D^{\frac{1}{2}} \mathbb{1} = 0 \iff f^T D^{\frac{1}{2}} D^{\frac{1}{2}} \mathbb{1} = f^T D \mathbb{1} = 0$$

So, when D^{-1} exists, there is a one-to-one mapping between the unnormalized and normalized formulations.