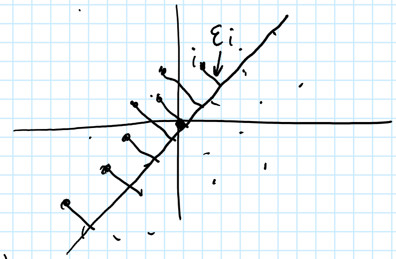
Lecture 14 notes

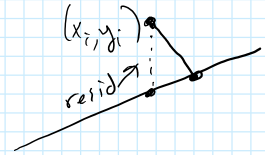
Wednesday, May 11, 2016 12:07 PM

$p=2$



$X_i \in \mathbb{R}^p$

(assume centered $\frac{1}{n}\sum_i X_i = 0$)



Principal Component Anaylsis (PCA)

w be s.t. subspace is span$\{w\}$ assume $\|w\|_2 = 1$

$\varepsilon_i^2 = \|X_i - w^T x_i w\|_2^2$ $(w^T x_i)w$ is projection of x_i onto the line

$= x_i^T x_i - 2x_i^T(w^T x_i)w + w^T(w^T x_i)^2 w = x_i^T x_i - 2(x_i^T w)^2 + (w^T x_i)^2 w^T w$

$= \|x_i\|_2^2 - (X_i^T w)^2$

$\min_w \sum_i \varepsilon_i^2 \Leftrightarrow \max_w \sum_i (X_i^T w)^2$

Sample Covariance for a coordinate j we know that $\frac{1}{n}\sum_{i=1}^n (x_{ij} - \underbrace{\bar{x}_j}_{0 \text{ by centered}})^2 = \hat{\sigma}_{jj}^2$

j,k then $\hat{\sigma}_{jk}^2 = \frac{1}{n}\sum_{i=1}^n x_{ij} x_{ik}$

Matrix $S = \begin{bmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{12}^2 & \cdots \\ \hat{\sigma}_{21}^2 & \ddots & \\ \vdots & & \end{bmatrix} = \frac{1}{n}\sum_{i=1}^n x_i x_i^T = \frac{1}{n} X^T X$

$\max_w \sum_i (x_i^T w)^2$ s.t. $\|w\|_2^2 = 1$

$L(\nu) = \sum_i (x_i^T w)^2 + \nu(w^T w - 1) = \sum_i w^T x_i x_i^T w + \nu(w^T w) - \nu$

$= w^T\left(\sum_i x_i x_i^T\right) w + \nu(w^T w) - \nu$

$= n\, w^T S w + \nu(w^T w) - \nu$

$\xrightarrow{\partial_w} 2nSw + 2\nu w \overset{set}{=} 0$ $\boxed{Sw = \left(-\frac{\nu}{n}\right) w}$ $\Rightarrow$ w is an eigenvector of S

$\triangleright$ w is the eigenvector w/ largest eigenvalue for S (λ)

$\max \sum_i (w^T x_i)^2 = n\, w^T S w = n\, w^T(\lambda w) = n\lambda w^T w = n\lambda$

Singular Value Decomposition (SVD)

of X is a triplet of matrices $\Sigma_{ii}^2 = \lambda_i$ the i^{th} eigenvalue

U : $n \times n$ is left eigenvectors $\qquad \Sigma_{ij} = 0 \quad i \neq j$

Σ : $n \times p$ are singular values $\qquad U^T U = I \quad V^T V = I$

V : $p \times p$ are right eigenvectors $\qquad X = U \Sigma V^T$

$$S = \frac{1}{n} X^T X = \frac{1}{n} (U \Sigma V^T)^T (U \Sigma V^T) = \frac{1}{n} V \Sigma^T U^T U \Sigma V^T = \frac{1}{n} V \Sigma^T \Sigma V^T$$

$= \frac{1}{n} V \Lambda V^T$ so v_i (columns of V) are eigenvectors of S.

1st Principal component: is (projection onto v_1 of x_i) $= (x_i^T v_1) v_1$

2nd " " : (" v_2 ") $= (x_i^T v_2) v_2$

Approximation of X $\qquad$ rank(X) = dim (row space of X) $\qquad |\sigma_1| \geq |\sigma_2| \geq \dots$

to approx X with $\hat{X} = U \hat{\Sigma} V^T$ where $\hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & 0 & & \\ & & \ddots & & & \\ 0 & & & \sigma_k & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix}$

distortion: $\sum_i \| x_i - \hat{x}_i \|_2^2$ is minimized for all rank k matrices.

High di embedding $\quad \Phi : \mathbb{R}^d \to \mathbb{R}^D$ then define $z_i = \Phi(x_i)$

$$S = \frac{1}{n} \sum_i z_i z_i^T = \frac{1}{n} Z^T Z \qquad Z^T Z = V \Lambda V^T \qquad Z = U \Sigma V^T$$

$$\underset{n \times n}{Z Z^T} = (U \Sigma V^T)(U \Sigma V^T)^T = U \Sigma V^T V \Sigma U^T = U \Lambda U^T$$

$$(Z^T U)_j = (V \Sigma U^T U)_j = (V \Sigma)_j = \sigma_j \cdot v_j$$

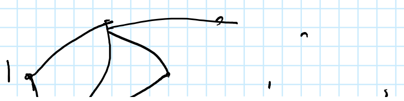
$$(Z Z^T)_{ij} = (\Phi(x_i))^T \Phi(x_j) = K(x_i, x_j)$$

Can compute U as eigenvectors of $K = Z Z^T$

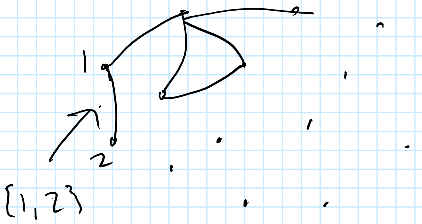
need $Z v_1, Z v_2, \dots, Z v_k \qquad Z v_j = U \Sigma V^T v_j = \sigma_j u_j = \sqrt{\lambda_j}\, u_j$

w/ λ_j eigenvalues.

Graph is a set of vertices and edges $\{1, \dots, n\} = V$,



edge set is list of unordered pairs of vertices (undirected)



edge set is list of unordered pairs of vertizes (undirected)

they can have weights ($w_{ij} \geq 0$ for vertices i,j)

Can write the $n \times n$ weight matrix as W ($W_{ij} = w_{ij}$)

One matrix is

combinatorial Laplacian $L = D - W$ w/ $D_{ii} = \sum_j w_{ij}$ and $D_{ij} = 0$ $i \neq j$.

normalized Laplacian $\tilde{L} = D^{-1/2} L D^{-1/2} = I - \underbrace{D^{-1/2} W D^{-1/2}}$

Laplacian eigenmaps

(1) Construct a graph $w_{ij} = k(x_i, x_j)$ $w_{ij} = e^{-t\|x_i - x_j\|_2^2}$

(2) Form normalized Laplacian, $\tilde{L}$, (can do this w/ L)

(3) Compute $\{\lambda_i, u_i\}$ spectral decomp. of $\tilde{L}$.

(4) Output vectors $(\sqrt{\lambda_1}\, u_{1j}, \ldots, \sqrt{\lambda_n}\, u_{nj})$ as the Laplacian eigenmap. $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.