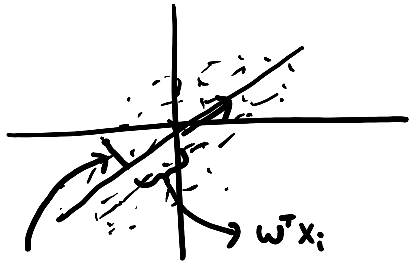
PCA

Thursday, May 4, 2017 10:13 AM

$x_i \in \mathbb{R}^p$ (assume centered $\frac{1}{n}\sum_i x_i = 0$)



$\varepsilon_i = \|x_i - w^T x_i w\|$

Principle Component Analysis:

Let w be s.t subspace is span$\{w\}$

assume $\|w\|_2 = 1$

Project x_i onto span$\{w\}$

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

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

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

Recall k-means : find m_k (center), z_i (assignment)

$\hat{x}_i = m_{z_i}$ distortion: $\sum_{i=1}^n \|x_i - \hat{x}_i\|_2^2$

In PCA $\hat{x}_i = w^T x_i w$

distortion: $\sum_i \|x_i - \hat{x}_i\|_2^2 = \sum_i \varepsilon_i^2$

$$\min_{\|w\|_2=1} \sum_i \varepsilon_i^2 \equiv \max_{\|w\|_2=1} \sum_i (x_i^T w)^2 = \max_{\|w\|_2=1} \|Xw\|_2^2$$

<u>Sample Covariance</u> $\hat{\sigma}_{jk}^2 = \frac{1}{n}\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$

Matrix $S = \begin{bmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{12}^2 & \cdots \\ \hat{\sigma}_{21}^2 & \hat{\sigma}_{22}^2 & \\ \vdots & & \ddots \end{bmatrix} = \frac{1}{n}\sum_{i=1}^n x_i x_i^T = \frac{1}{n} X^T X$

$$\max_w \underbrace{\sum_i (x_i^T w)^2}_{} \quad \text{s.t } \|w\|_2^2 = 1$$

$$\max_{w} \underbrace{\sum_i (x_i' w)} \quad \text{s.t. } \|w\|_2^2 = 1$$

$$\sum_i w^T x_i x_i^T w = w^T \left(\sum_i x_i x_i^T \right) w = n\, w^T S w$$

$$\equiv \max_{w} w^T S w \quad \text{s.t. } \|w\|_2^2 = 1$$

$$\equiv \max_{w} w^T S w \quad \text{s.t. } \|w\|_2^2 \le 1 \quad (\text{non convex!})$$

$$L(w, \nu) = w^T S w + \nu (w^T w - 1)$$

$$\xrightarrow{\partial_w} 2Sw + 2\nu w \overset{\text{set}}{=} 0$$

$$Sw = (-\nu) w \implies$$ solution to PCA is eigenvector of S

w is eigenvector w/ eigenvalue ξ

$$w^T S w = w^T (\xi w) = \xi$$

▸ for $d = 1$ (where d is the reduction dimension)

Singular Value Decomp

X has SVD

U: $n \times n$ left eigenvect.

Σ: $n \times p$ singular values

V: $p \times p$ right eigenvect.

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \quad \sigma_p \\ & 0 & \end{bmatrix}$$

$$X = U \Sigma V^T \qquad U^T U = I \qquad V^T V = I$$

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

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

$$= \frac{1}{n} V \Sigma^T \Sigma V^T \qquad \Sigma^T \Sigma = \begin{bmatrix} \sigma_1 & & & & 0 \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_p & \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \\ & 0 & \end{bmatrix}$$

$$= \frac{1}{n} V (n \Xi) V^T \qquad n \Xi = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \ \sigma_p^2 \end{bmatrix}$$

$$= V \Xi V^T$$

SVD of $X \longleftrightarrow$ Spectrum of S

$X = U \Sigma V^T \qquad XV = U\Sigma \rightarrow$ get U

def 1st Principle Component: proj. onto v_1 of x_i

$$= (x_i^T v_1) v_1$$

2nd PC: proj onto v_2 of x_i $= (x_i^T v_2) v_2$

Approx X construct "low rank" version of X

$$\hat{X} = U \hat{\Sigma} V^T \quad w/ \quad \hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_{d_0} & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \ldots$

minimized distortion: $\sum_i \|x_i - \hat{x}_i\|_2^2$