Lecture November 4

project 2

NN : regression.

in p1

$$f(x) \simeq P_N(x) = \sum_{j=0}^{N} \beta_j x^j$$

$$\rightarrow \quad X = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots \\ 1 & \vdots & & \\ \vdots & \vdots & & \\ 1 & x_{n-1} & \cdots & x_{n-1}^N \end{bmatrix}$$

For NNs, how should X look like

$$X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \leftarrow \text{centered}$$

or

$$X = \begin{bmatrix} 1 & x_0 & \cdots & x_0 \\ 1 & & & \vdots \\ \vdots & & & \vdots \\ 1 & & & x_{n-1}^{p-1} \end{bmatrix}$$

$$y = f(x) + \varepsilon$$

$$y_i = f(x_i) \simeq NN(x_i)$$

PCA

$$C[X] = \begin{bmatrix} var[x_0] & - & - \\ cov[x_1, x_0] & - & - \\ cov[x_2, x_0] & - & - \\ \vdots & & \\ cov[x_{p-1}, x_0] & - & - \\ & & cov[x_0, x_{p-1}] \\ & & \vdots \\ & & var[x_{p-1}] \end{bmatrix}$$

$$cov[x_i, x_j] = \frac{1}{n} \sum_{l=0}^{n-1} (x_{li} - \mu_{x_i})$$

$$\times (x_{ej} - \mu_{x_j})$$

$$X = \begin{bmatrix} | & | & & | \\ x_0 & x_1 & \cdots & x_{p-1} \\ | & | & & | \end{bmatrix}$$

$$\in \mathbb{R}^{n \times p}$$

$$\mathrm{var}[x_i] = \frac{1}{n} \sum_{j=0}^{n-1} (x_{ji} - \mu_{x_i})^2$$

$$\mu_{x_i} = \frac{1}{N} \sum_{j=0}^{n-1} x_{ji}$$

Assumed that

$$C[y] = S\, C[x]\, S^T = \mathrm{Diag}(\lambda)$$

$$\frac{1}{n} X^T X$$

$$S S^T = S^T S = \mathbb{I}$$

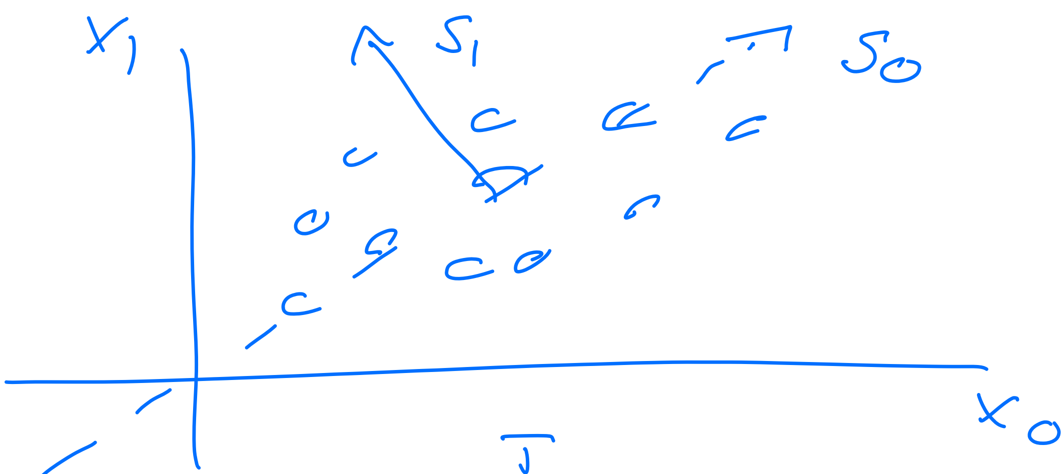
$$\mathrm{Diag}(\lambda) = \begin{bmatrix} \mathrm{var}[y_0] & 0 & \cdots & 0 \\ 0 & \mathrm{var}[y_1] & & \\ \vdots & & & \end{bmatrix}$$

$$\begin{bmatrix} 0 & & \mathrm{var}[y_{p-1}] \end{bmatrix}$$

$$\lambda_i = \mathrm{var}[y_i]$$

$\mathrm{Cov}[y] = 0$, the variables y_i are uncorrelated.

$$y_i^T y_j = \delta_{ij}$$



$$s_1^T s_0 = 0$$

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

$$X^T X = n\, C[x] = n\, C_x$$

$$(X^T X) v_i = \sigma_i^2 v_i$$

$$V^T = \begin{bmatrix} | & | & & | \\ v_0 & v_1 & \cdots & v_{p-1} \\ | & | & & | \end{bmatrix}$$

$$S C_x S^T = C_y = \mathrm{Diag}(\lambda)$$

$$\lambda_i = \mathrm{var}[y_i] \propto \sigma_i^2$$

the vectors s_i are proportional to v_i

PCA Theorem (Statistical view)

Define

$$y_i = s_i^T x$$

$$s_i^T s_i$$

we require that the variance of y_i is so

" +

that

$$\mathrm{var}[y_0] \geq \mathrm{var}[y_1] \cdots \geq \mathrm{var}[y_d] > 0$$

Total dim is D

we want a $-d-$ which satisfies $d < D$

$$\hat{S}_i = \arg\max_{S_0 \in \mathbb{R}^D} \mathrm{var}[S_0^T x]$$

optimal

subject to $S_0^T S_0 = 1$

Theorem

The first d-principal components of a zero mean value multivariate variable x, denoted by $y_i = 0, 1, 2, \ldots d-1$, are given by

$$y_i = S_i X$$

where S_i are orthonormal vector of $C_x = \frac{1}{n} X^T X$ associated with its d-largest eigenvalues

$$\lambda_i = var[y_i]$$

$$\hat{S}_0 = \boxed{\max \; S_0^T C_x S_0}$$

$$\text{s.t.} \quad S_0^T S_0 = \underline{1}$$

Lagrange multiplier

$$\mathcal{L} = S_0^T C_x S_0 + \lambda_0 (1 - S_0^T S_0)$$

$$\left(\begin{array}{c} f(x) + \lambda g(x) \\ g(x) = 0 \end{array} \right)$$

Take derivatives of $\mathcal{L}$ wrt S_0 and λ_0

$$\boxed{C_x s_0 = \lambda_0 s_0}$$

$$s_0^T s_0 = \underline{1}$$

$\rightarrow$ eigenvalue problem

s_0 is an eigenvector of C_x with eigvalue λ_0. Max problem

$\Rightarrow$ λ_0 is largest eigenvalue of C_x

$$\lambda_0 = var[y_0]$$

To find the second one

$$s_1^T s_0 = 0$$

$$\mathcal{L} = s_1^T C_x s_1 + \lambda_1 (1 - s_1^T s_1) + \gamma s_1^T s_0$$

derivatives wrt s_1, λ_1, γ

$$C_x s_1 + \frac{\gamma}{2} s_0 = \lambda_1 s_1$$

$$s_1^T s_1 = 1 \quad \text{and} \quad s_1^T s_0 = 0$$

$\times s_0^T$

$$s_0^T C_x s_1 + \frac{\gamma}{2} \underbrace{s_0^T s_0}_{=1} = \lambda_1 s_0^T s_1$$

$$\lambda_0 s_0^T \quad \lambda_1 s_1$$

$$\gamma = 2(\lambda_1 - \lambda_0)\, s_0^T s_1 = 0$$

$$\Rightarrow \quad C_x s_1 = \lambda_1 s_1$$

$$s_1^T C_x s_1 = \lambda_1 = \mathrm{var}[y_1]$$

continue by induction

$$\lambda_i = \mathrm{var}[y_i]$$

$$s_i^T s_j = \delta_{ij}$$

We keep only those s_i which have a

variance $\text{var}[y_i] = \lambda_i$
(eigenvalues of C_x)
smaller than a
prefixed value ε