FYS5429/9429, lecture March 20, 2025

FYS5429/9429 March 20

PCA and AEs

assume linear AE

AE:

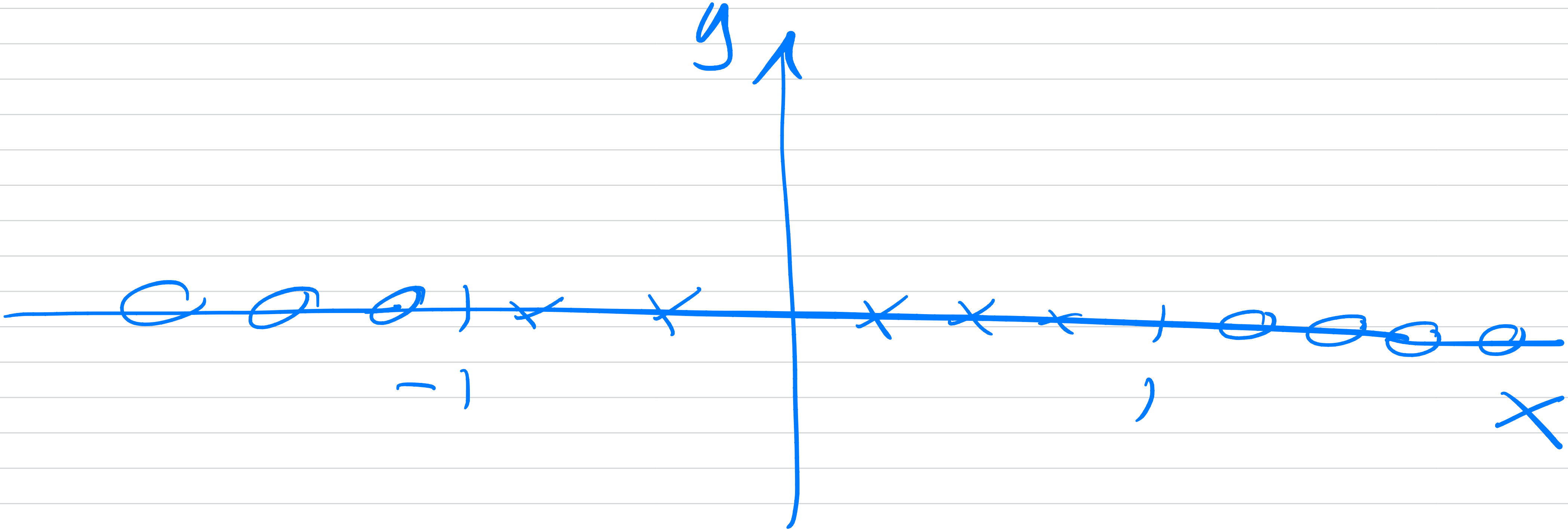
$$\underset{W\,\tilde{V}}{\arg\min} \frac{1}{2}\sum_i (y_i - \tilde{y}_i)^2$$

$$h = Wy \qquad g = \tilde{y} = \tilde{V}h$$

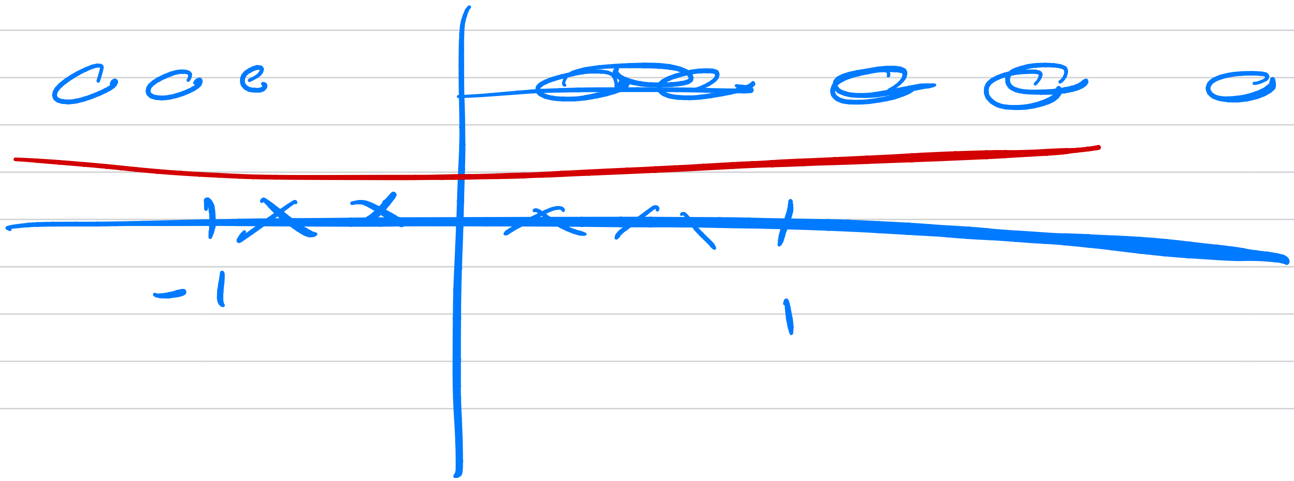
$$\tilde{y} = \tilde{V}Wy$$

$$\underset{W\,\tilde{V}}{\arg\min} \frac{1}{2}\|(y - \tilde{V}Wy)\|_2^2$$

non-$\boxed{\text{linear}}$ (kernel) PCA



kernel: x^2



$$\operatorname{cov}[\vec{x}_i, \vec{x}_j] = \frac{1}{m} \sum_k (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

$$\vec{x}_i = \begin{bmatrix} x_{0i} \\ x_{1i} \\ \vdots \\ x_{m-1\,i} \end{bmatrix} \qquad \bar{x}_i = \frac{1}{m} \sum_{j=0}^{m-1} x_{ji}$$

$$\tilde{x}_{ki} = x_{ki} - \bar{x}_i \longrightarrow x_{ki}$$

$$\operatorname{cov}[\vec{x}_i, \vec{x}_j] = \frac{1}{m} \vec{x}_i^{\,T} \vec{x}_j$$

$$\vec{x}_i \vec{x}_j^{\,T} = \text{Matrix}$$

$$X = \begin{bmatrix} | & | \\ \vec{x}_0 & \vec{x}_1 \\ | & | \end{bmatrix}$$

$$\operatorname{cov}[X] = \begin{bmatrix} \sigma_0^2 & \operatorname{cov}[\vec{x}_0, \vec{x}_1] \\ \operatorname{cov}[\vec{x}_1, \vec{x}_0] & \sigma_1^2 \end{bmatrix}$$

in general

$$X = \begin{bmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_{p-1} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times p}$$

Covariance matrix

$$\mathrm{Cov}[X] = \frac{1}{n} X^T X \in \mathbb{R}^{p \times p}$$

$$\tilde{x}_{ki} = x_{ki} - \bar{x}_i \Rightarrow x_{ki}$$

$$\mathrm{var}[\vec{x}_i] = \frac{1}{n} \sum_k x_{ki}^2$$

$$\mathrm{corr}[\vec{x}_i \, \vec{x}_j] = \frac{\mathrm{cov}[\vec{x}_i \, \vec{x}_j]}{\sqrt{\mathrm{var}[\vec{x}_i]\,\mathrm{var}[\vec{x}_j]}}$$

$$\mathrm{corr}[X] = \begin{bmatrix} 1 & \mathrm{corr}[\vec{x}_0 \vec{x}_1] & \cdots & \mathrm{corr}[\vec{x}_0 \vec{x}_{p-1}] \\ \vdots & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathrm{corr}[\vec{x}_{p-1} \vec{x}_0] & \cdots & & 1 \end{bmatrix}$$

$\vec{x}_n$ are stochastic variables

and if all $\vec{x}_n$ are i.i.d.

independent

identically

distributed

$$\Rightarrow \quad cov[\vec{x}_i\ \vec{x}_j] = 0$$
$$i \neq j$$

$$Cov[X] = \begin{bmatrix} \sigma_0^2 & & & \\ & \sigma_1^2 & & 0 \\ & & \ddots & \\ 0 & & & \sigma_{p-1}^2 \end{bmatrix}$$

$$\mathrm{cov}[X] = \frac{1}{n} X^{T} X \in \mathbb{R}^{p \times p}$$

To find eigenvalues

$$S S^{T} = S^{T} S = \mathbb{1}_{p \times p}$$

aside

$$\bar{X} = \mathbb{E}[X] = \langle X \rangle = \int_{x \in D} P_X(x)\, x\, dx$$

$$\left(\sum_{x_i \in D} P_X(x_i)\, x_i \right.$$

we deal with sample

$$\text{mean} = \mu_X = \frac{1}{n} \sum_i x_i \neq \bar{X}$$

$$\operatorname{cov}[X] = \mathbb{E}[X^T X] = \frac{1}{n} X X^T$$

$$\underline{S^T X^T X S}$$

aside

$$\left[\begin{array}{l} Av = \lambda v \\ S^T A S = D \end{array}\right.$$

$$= \begin{bmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{n-1} \end{bmatrix}$$

$$\operatorname{cov}[X] \rightarrow$$

$$\mathbb{E}[S^T X^T X S] =$$

$$S^T \mathbb{E}[X^T X] S =$$

$$\operatorname{cov}[Y]$$

$$\frac{1}{n} Y^T Y = \frac{1}{n} S^T X^T X S$$

$$= \begin{bmatrix} \sigma_0^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sigma_{p-1}^2 \end{bmatrix}$$

Link with SVD

$$X = U \Sigma V^T \in \mathbb{R}^{n \times p}$$

$$U U^T = U^T U = \mathbb{1}_{n \times n}$$

$$V V^T = V^T V = \mathbb{1}_{p \times p}$$

$$\Sigma = \begin{bmatrix} \lambda_0 & & & 0 \\ & \lambda_1 & & \\ & & \ddots & \\ 0 & & & \lambda_{p-1} \\ & & & 0 \end{bmatrix} \in \mathbb{R}^{n \times p}$$

$$\mathrm{cov}[X] = \frac{1}{n} X^T X$$

$$= \frac{1}{n} \underbrace{V \Sigma^T}_{p \times n} \underbrace{U^T U}_{I_{n \times n}} \underbrace{\Sigma}_{n \times p} V^T$$

$$= \frac{V \Sigma^T \Sigma V^T}{n} = V \Sigma^2 V^T$$

$$\lambda_0^2 = \sigma_0^2$$

$$= \frac{V}{n} \begin{bmatrix} \lambda_0^2 & & & \\ & \lambda_1^2 & & \\ & & \ddots & \\ & & & \lambda_{p-1}^2 \end{bmatrix} V^T$$

$$\lambda_0 \geq \lambda_1 \geq \lambda_2 \ldots \geq \lambda_{p-1} \geq 0$$

multiply with V from the right

$$\frac{1}{n}(X^TX)V = \frac{1}{n} V \Sigma^2 \underbrace{V^T V}_{\uparrow\, p\times p}$$

$$V = \begin{bmatrix} | & | & & | \\ \vec{v}_0 & \vec{v}_1 & \cdots & \vec{v}_{p-1} \\ | & & & | \end{bmatrix}$$

$$v_i^T v_j = \delta_{ij}$$

$$\frac{1}{n}(X^TX)v_i = \frac{1}{n} v_i \lambda_i^2 = v_i \sigma_e^2$$

PCA approx:

Example: keep 95% of total variance

$$\sum_{i=0}^{p-1} \sigma_i^2$$

truncate to keep only those singular values which 95% of total variance

Statistical PCA

Reconstruction error: want to minimize

$$J(y) = \frac{1}{n} \sum_i (y_i - \tilde{y}_i)^2$$

$$\tilde{y} = \tilde{V} W y$$

$$J(y) = \frac{1}{n} \sum_i (y_i - (\tilde{V} W)\tilde{y}_i)^2$$

we want to find

$$\tilde{V}W = \vec{s}_0 \vec{s}_0^{\,T}$$ — outer product

$$(\vec{s}_0^{\,T} \vec{s}_0)$$

inner product

$$\underset{s_0}{\operatorname{argmin}} \frac{1}{n} \left\| \left(\vec{y} - \vec{s}_0 \vec{s}_0^{\,T} \vec{y} \right) \right\|_2^2$$

$$S = \begin{bmatrix} | & | & | & \\ \vec{s}_0 & \vec{s}_1 & \vec{s}_2 & \cdots \\ | & | & | & \end{bmatrix}$$

$$\tilde{\nabla} W \propto \vec{S}_0 \vec{S}_0^T$$

$$\vec{S}_0^T \vec{S}_0 = 1$$

Define a Lagrangian

$$\vec{S}_0^T \operatorname{cov}[\vec{x}] \vec{S}_0 + \lambda_0 (1 - \vec{S}_0^T \vec{S}_0)$$

optimize wrt to $\vec{S}_0$ and λ_0

wrt $\vec{S}_0^T$ $\qquad \vec{S}_0^T$

$$\Downarrow$$

$$\operatorname{cov}[\vec{x}] \vec{S}_0 = \lambda_0 \vec{S}_0$$

$$\vec{s}_0^{\,T}\, cov[X]\, \vec{s}_0 = \lambda_0$$

eigenvalue on $cov[X]$

$\lambda_0 = \sigma_0^2$ (largest eigenvalue)

$\vec{s}_0$: first principal component

Next is $\vec{s}_1$

$$\vec{s}_1^{\,T} \vec{s}_0 = 0$$

$$\mathcal{L} = \vec{S}_1^T \operatorname{cov}[X]\, \vec{S}_1 +$$

$$\lambda_1 (1 - \vec{S}_1^T \vec{S}_1) + \gamma\, \vec{S}_1^T S_0$$

Take derivatives wrt $\vec{S}_1$, λ_1 and γ

$\underline{S_1}$

$$\operatorname{cov}[X]\, \vec{S}_1 - \gamma\, \vec{S}_0 / 2 = \lambda \vec{S}_1$$

multiply ~~from~~ left with S_0'

$$\Big(\underbrace{\vec{s}_0^{\,T} \operatorname{cov}[x]}_{\lambda_0 \vec{s}_0^{\,T}} \Big) \vec{s}_1 + \gamma/2\, \vec{s}_0^{\,T} \vec{s}_0$$

$$\lambda_0 \underbrace{\vec{s}_0^{\,T} \vec{s}_1}_{=0} = \lambda_1 \vec{s}_0^{\,T} \vec{s}_1$$

$$\gamma = 2 [\lambda_1 - \lambda_0] \underbrace{\vec{s}_0^{\,T} \vec{s}_1}_{=0}$$

$$\gamma = 0$$

$$\vec{s}_1^{\,T} \operatorname{cov}[x] \vec{s}_1 = \lambda_1 = \operatorname{var}[y_1]$$

second

by induction

$$\mathcal{L} = \vec{s}_i^{\,T} \operatorname{cov}[X] \vec{s}_i \qquad \vec{s}_i \vec{s}_i^{\,T}$$

$$+ \lambda_i \left(1 - \vec{s}_i^{\,T} \vec{s}_i\right)$$

$$+ \sum_{j=0}^{i-1} \gamma_j \vec{s}_i^{\,T} \vec{s}_j \qquad \tilde{v} \, W$$

AE :

$$\underset{V\,W}{\operatorname{argmin}} \frac{1}{n} \sum \left(y_i - g(\tilde{v}) f(w) \times \tilde{y}_i\right)$$