FYS5429/9429 Lecture March 13, 2025

FYS5429/9429 March 13

input Encoder Latent feature representation

$y_0 \quad h_0$

$y_1 \quad h_1$

$\vdots \quad \vdots$

$y_{\boxed{n-1}} \quad h_{n-1} \quad \Longrightarrow$

Decoder output

$\Longrightarrow \quad s_0 \quad \tilde{y}_0$

$\vdots \quad \vdots$

$s_{l-1} \quad \tilde{y}_{\boxed{n-1}}$

$$h = f(W, y)$$

$$\tilde{y} = g(V, h)$$

Define a metric for success

$$MSE = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

optimization

$$\underset{W,V}{\operatorname{argmin}} \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

Link with PCA and AEs

$$\boxed{h = Wy \qquad g = V \cdot h}$$ linear

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

$$\underset{W,\,V}{\arg\min} \ \frac{1}{n} \sum_{i=0}^{n-1} \left[(1 - VW)\, y \right]^2$$

AEs are in general not linear

Design matrix

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

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0\,p-1} \\ x_{10} & x_{11} & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ x_{n-1\,0} & x_{n-1\,1} & \cdots & x_{n-1\,p-1} \end{bmatrix}$$

data entries

Features

OLS

$$\tilde{y} = X\hat{e} \; ; \; \underbrace{\hat{e}}_{\mathbb{R}^p} = \frac{1}{\underbrace{\boxed{X^T X}}_{\mathbb{R}^{p\times p}}} \underbrace{\boxed{X^T}}_{\mathbb{R}^{p\times n}} \underbrace{y}_{\text{output}}^{\mathbb{R}^n}$$

$$\tilde{y} = \underbrace{X \frac{1}{X^T X} X^T}_{\hat{A}} y \qquad \hat{A}^2 = \hat{A}$$

$$\|y - \tilde{y}\|_2^2 \qquad \left(\|v\|_2^2 = \sum_i v_i^2 \right)$$

$$y - \tilde{y} = (\mathbb{1} - A)\, y$$

$$= (\mathbb{1} - A^2)\, y$$

$$A^2 = V \cdot W$$ from a linear AE

no reduction of dimension

SVD (and dim-reduction)

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

$$X = U D V^T$$

$$U U^T = U^T U = \mathbb{1} = V V^T = V^T V$$

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

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

$$D = \begin{bmatrix} \lambda_0 & & & 0 \\ & \lambda_1 & & \\ 0 & & \ddots & \\ & & & \lambda_{p-1} \\ & & & 0 \end{bmatrix}$$

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

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

$$\ldots 0$$

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad D \in \mathbb{R}^{3 \times 2}$$

OLS

$$\hat{y} = X\hat{\theta} = X \frac{1}{X^T X} X^T y$$

$$X = U D V^T$$

$$\hat{y} = U D V^T \left[\frac{1}{V D^T U^T U D V^T} \right] \times V D^T U^T y$$

$$= U D V^T \frac{1}{V D^T D V^T} V D^T U^T y$$

$$D^T D = \begin{bmatrix} \lambda_0^2 & & \\ & \ddots & \\ & & \lambda_{p-1}^2 \end{bmatrix}$$

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

non-zero and invertible

V is also invertible

if A and B are square and invertible matrices

$$\frac{1}{AB} = B^{-1} \cdot A^{-1}$$

$$V = \begin{bmatrix} | & | & | & & | \\ v_0 & v_1 & v_2 & \cdots & v_{p-1} \\ | & | & | & & | \end{bmatrix}$$

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

$$V = (V^T)^{-1}$$

$$\tilde{y} = \underbrace{UD}_{\mathbb{R}^{n \times n}} \frac{1}{\underbrace{D^T D}_{\mathbb{R}^{p \times p}}} D^T U^T y$$

$n > p$

$$= \left[\sum_{i=0}^{p-1} u_i u_i^T \right] y$$

PCA : we want a dim $m < p$

want to minimize the reconstruction (AEs)

$$\| (I - VW) y \|_2^2 \leq \varepsilon$$

$$\| (y - \tilde{y}) \|_2^2 \leq \varepsilon$$

show that a linear AE ($V W$) corresponds to linear PCA

SVD

$$V W_y = \overline{\underbrace{U U^T}} y$$

(which only first p-components)

Covariance

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

$$X = \begin{bmatrix} | & | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_{p-1} \\ | & | & | & & | \end{bmatrix}$$

$$\operatorname{cov}[x_i, x_j] \ (\text{sample}) = \frac{1}{n} \sum_k (x_{ki} - \bar{x}_i) \times (x_{kj} - \bar{x}_j)$$

$$\vec{x}_i = \begin{bmatrix} x_{0i} \\ x_{1i} \\ \vdots \\ x_{n-1\,i} \end{bmatrix}$$

$$\bar{x}_i = \frac{1}{n} \sum_k x_{ki}$$

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

$$X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \\ x_{20} & x_{21} \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{x}_0 & \vec{x}_1 \\ | & | \end{bmatrix}$$

$$\mathrm{cov}[\vec{x}_0, \vec{x}_1] = \frac{1}{3} \sum_k x_{k0} x_{k1}$$

$$= \frac{1}{3}(x_{00}x_{01} + x_{10}x_{11} + x_{20}x_{21})$$

$$= \frac{1}{3}\begin{bmatrix} x_{00} & x_{10} & x_{20} \end{bmatrix}\begin{bmatrix} x_{01} \\ x_{11} \\ x_{21} \end{bmatrix} = \frac{1}{3}\vec{x}_0^{\,T}\vec{x}_1$$

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

our example

$$\operatorname{cov}[X] = \frac{1}{n} \begin{bmatrix} \vec{x}_0^T \vec{x}_0 & \vec{x}_0^T \vec{x}_1 \\ \vec{x}_1^T \vec{x}_0 & \vec{x}_1^T \vec{x}_1 \end{bmatrix}$$

$$\frac{1}{n} \vec{x}_0^T \vec{x}_0 = \frac{1}{n} \sum_k x_{k0}^2$$

$$= \frac{1}{n} \sum_k \left(x_{k0} - \bar{x}_0 \right)^2$$

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

$$\sigma_i^2 = \text{variance}$$

$$= \frac{\vec{x}_i^T \vec{x}_i}{n} = \mathrm{var}[x_i]$$

correlation matrix

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

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