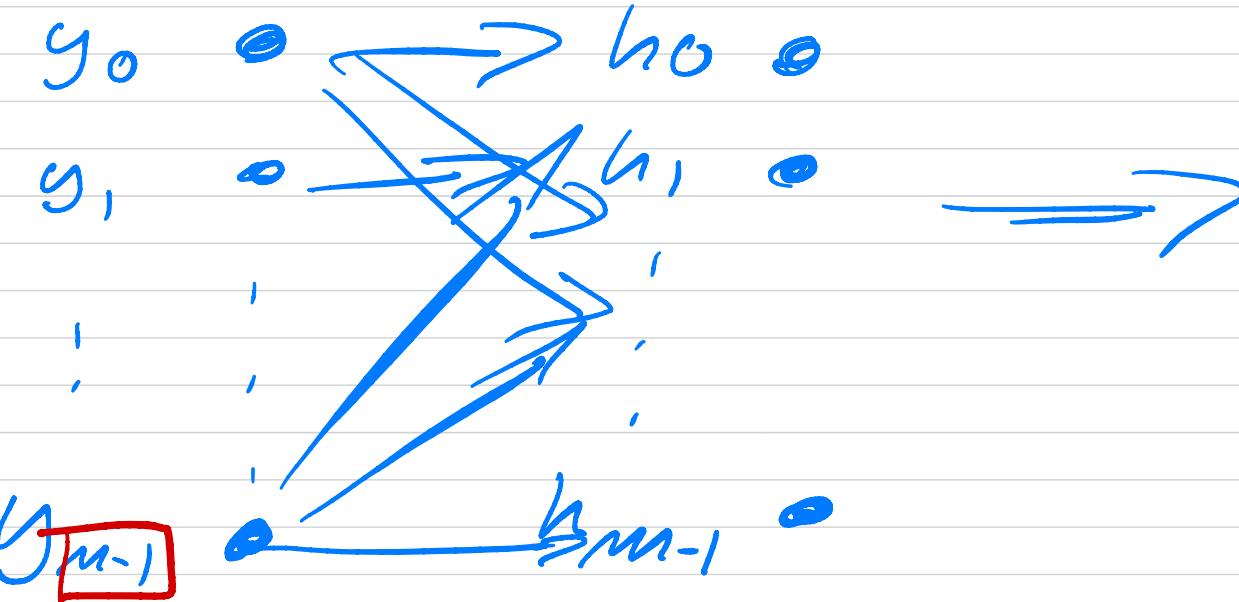


FYS5429/9429 Lecture

March 13, 2025

FYS5429/9429 March 13

input

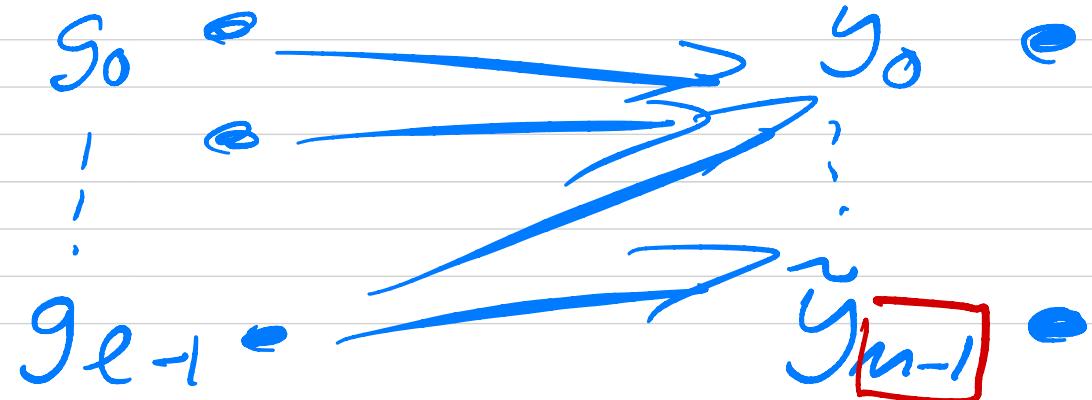


Encoder

Latent
feature
representation

Decoder

output



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

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

Define a metric for success

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

optimization

$$\arg \min_{w, v} \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2$$

Link with PCA and AEs

$$h = Wy \quad g = V \cdot h \quad \text{linear}$$

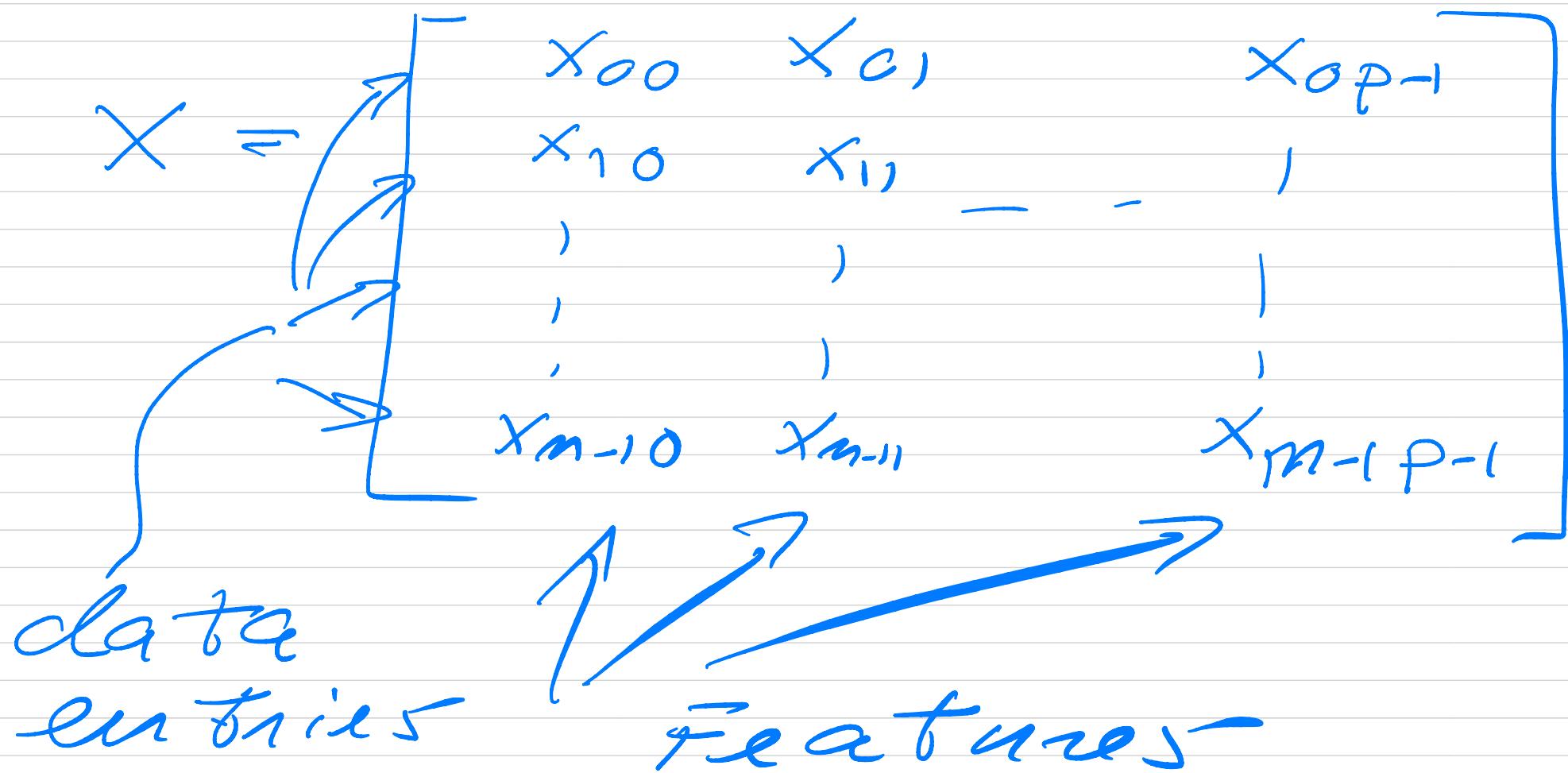
$$\tilde{y} = V \cdot W \cdot y$$

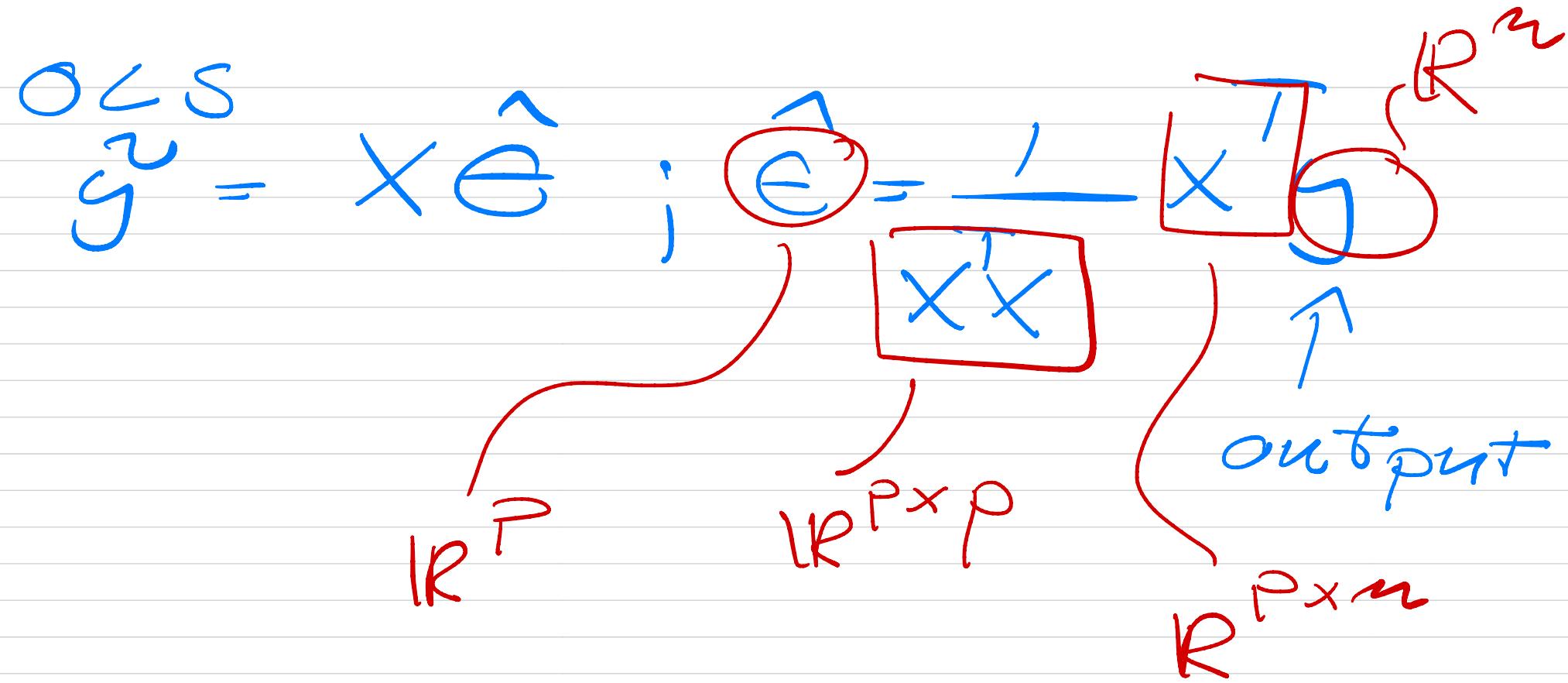
$$\underset{w, v}{\arg \min} \frac{1}{n} \sum_{i=0}^{n-1} [(\tilde{y}_i - Vw) \cdot y_i]^2$$

AEs are in general not linear

Design matrix

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





$$\hat{g}^s = \hat{x}^s ; \quad \hat{e} = \frac{1}{\hat{x}^T \hat{x}} \hat{x}^T \hat{g}^s$$

$$\hat{A}^s = \hat{A}$$

$$\|y - \hat{y}\|_2^2$$

$$\left(\|v\|_2 = \sqrt{\sum_i v_i^2} \right)$$

$$y - \hat{y} = (I - A)y$$

$$= (I - A^2)y$$

$$A^2 = V \cdot W$$

no reduction
of dimension

from a
linear
 $A \in$

SVD (and dim-reduction)

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

$$X = UDV^T$$

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

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

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

$$D = \begin{bmatrix} -\gamma_0 \gamma_1 & 0 \\ C & -\gamma_{p-1} \end{bmatrix}$$

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

$$\gamma_0 > \gamma_1 > \gamma_2 > \dots > \gamma_{p-1} > 0,$$

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

$\mathcal{O} \subset S$

$$\hat{y} = \hat{x}^T = x \frac{1}{x^T x} x^T y$$

$$x = u D v^T$$

$$\hat{y} = u D v^T \left[\frac{1}{v^T u^T u D v^T} \right] x v^T u^T y$$

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

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

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

non-zero and invertible

V is also invertible

if A and B are square
and invertible matrices

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

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

$$N_i^T v_j = \delta_{i,j}$$

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

$$\tilde{y} = uD \frac{'}{\underbrace{D^T D}} D^T u^T y$$

$u \in \mathbb{R}^{n \times n}$
 $D \in \mathbb{R}^{P \times P}$
 $n > P$

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

PCA : we want a dimension

$$m < p$$

want to minimize the reconstruction (AE_R)

$$\|(\mathbf{I} - \mathbf{V}\mathbf{W})\mathbf{y}\|_2^2 \leq \varepsilon$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 \leq \varepsilon$$

Show that a linear AE
($\mathbf{V}\mathbf{W}$) corresponds to

linear PCA

SVD

$$\sqrt{\lambda} \mathbf{w}_y = \underbrace{\mathbf{u} \mathbf{u}^T}_{\text{COLUMNS ARE BY FIRST}} \mathbf{y}$$

(columns are by first
P-components)

Covariance

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

$$X = \begin{bmatrix} \bar{x} \\ x_0 & x_1 & \dots & x_{p-1} \\ | & | & \uparrow & | \end{bmatrix}$$

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

(sample)

$$\bar{x}_i = \begin{bmatrix} x_{0i} \\ x_{1i} \\ \vdots \\ x_{(n-1)i} \end{bmatrix}$$

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

$$\tilde{x}_{k_1} = x_{k_1} - \bar{x}_1 \Rightarrow x_{k_1}'$$

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

$$CCV[\tilde{x}_0, \tilde{x}_1] = \frac{1}{3} \sum x_{k_0} x_{k_1}$$

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

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

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

our example

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

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

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

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

σ^2 = variance

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

Correlation matrix

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

$\text{con}[\vec{x}] =$

$$\left[\begin{matrix} 1 & \text{con}[\vec{x_0x_1}] & \dots & \text{con}[\vec{x_0x_{p-1}}] \\ & | & & | \\ & & \ddots & \\ & | & & | \\ & & & 1 \\ \text{con}[\vec{x_{p-1}x_0}] & & & & & | \\ & & & & & & | \end{matrix} \right]$$