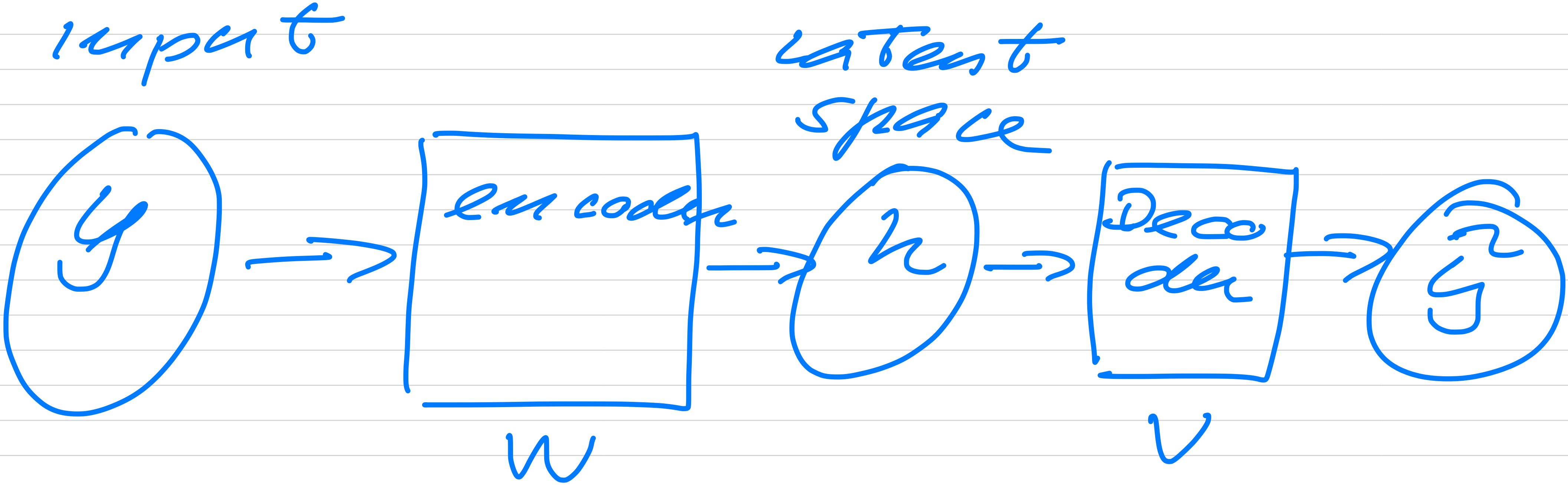
FYS-STK3155/4155, lecture November 24, 2025

FYS-STK3155/4155 November 24

Autoencoders



Linear model

$$h = W \cdot y = f(W, y) \qquad \tilde{y} = g(V, h) = V \cdot h$$

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

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

optimization

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

Linear approx

$$\frac{1}{n} \sum_i \left[(1 - V \cdot W) y_i \right]^2$$

AE are in general not linear.

How do we link AE (linear) with (linear) PCA?

OLS with a design matrix $X \in \mathbb{R}^{n \times p}$

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

$$A^2 = A \; ; \quad \hat{y} = A^2 y = A y$$

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

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

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

linear model

$$A^2 = V \cdot W$$ — linear AE

No reduction of dimension

SVD

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

$$X = U D V^T$$

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

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

$$U \in \mathbb{R}^{n \times n} \qquad V \in \mathbb{R}^{p \times p}$$

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

$$\tilde{y} = \left(\sum_{i=0}^{p-1} u_i u_i^T \right) y$$

Linear PCA and linear AE

We need the covariance in order to set up PCA

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

$$\vec{x}_i^T = [x_{0i}, x_{1i}, \ldots x_{n-1\,i}]$$

$$\mathrm{cov}[\vec{x}_0, \vec{x}_1] = \,?$$

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

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

$(\tilde{x}_{ki} \rightarrow x_{ki})$ our standard scaling

$$\mathrm{cov}[\vec{x}_i, \vec{x}_j] = \frac{1}{n} \sum_k (\tilde{x}_{ki}) \tilde{x}_{kj}$$

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

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

$$= \frac{1}{3} \begin{bmatrix} x_{00} & x_{10} & x_{20} \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{11} \\ x_{21} \end{bmatrix}$$

$$= \frac{1}{3} \vec{x}_0^{\,T} \vec{x}_1$$

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

an example

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

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

$$\text{variance} = \frac{1}{n} \sum_k (x_{k0} - \bar{x}_0)^2 = \sigma_0^2$$

$$\mathrm{cov}[X] =$$

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

correlation matrix

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

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

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

$$S^T S = S S^T = 1$$

$$S \frac{1}{n} X^T X S^T = S\,\mathbb{E}[X^T X]\, S^T$$

non-stochastic

$$= \text{Cov}[\bar{g}]$$

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

$$\begin{bmatrix} \tilde{\sigma}_0^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \tilde{\sigma}_{p-1}^2 \end{bmatrix}$$

variance

Back to SVD

$$X = UDV^T$$

$$\frac{1}{n} X^T X = \frac{1}{n} V D^T \underbrace{U^T U}_{I} D V^T$$

$$= \frac{1}{n} \underbrace{V D^T D V^T}_{p \times p}$$

$$\cancel{\frac{1}{n}} X^T X = \cancel{\frac{1}{n}} V D^2 V^T$$

$$\overline{X^T X} = \overline{V D^2 V^T} V$$

$$\left(\overline{X^T X}\right) V = V D^2$$

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

$$\left(\overline{X^T X}\right) \vec{v}_i = \vec{v}_i \overline{\sigma_i^2}$$

The singular values $\frac{\sigma_i^2}{n}$ are the variances of the covariance matrix

$$\sigma_0^2 \geq \sigma_1^2 \geq \sigma_2^2 \cdots \geq \sigma_{p-1}^2 \geq 0$$

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

$$\sum_{i=0}^{d} \sigma_i^2 \leq \Lambda$$

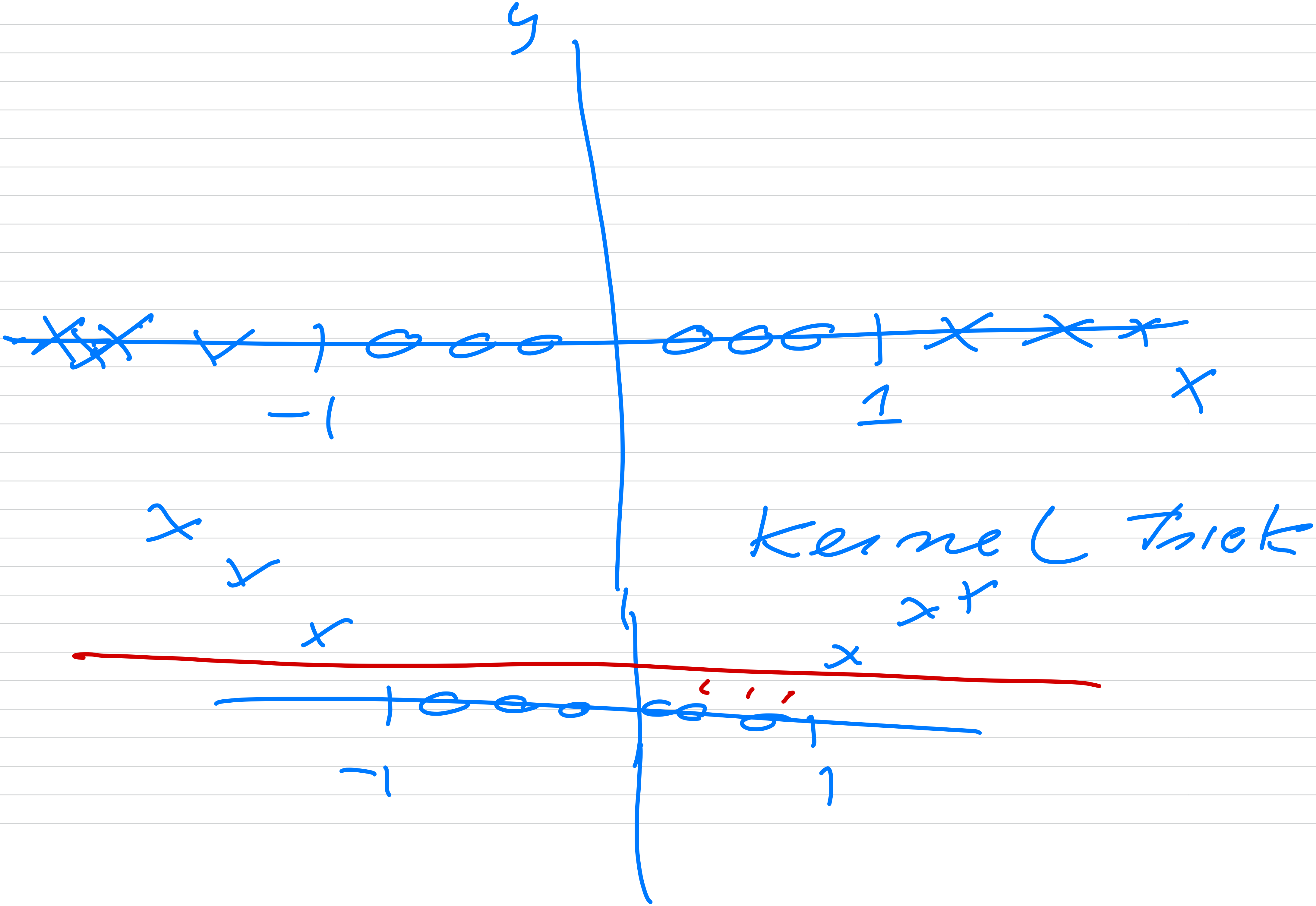
$$\frac{1}{n}\sum_{i=0}^{n-1}\left[(1 - VW)y_i\right]^2$$

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

in an SVD analysis
(Basis of PCA)

$$\tilde{y} = \left(\sum_{i=0}^{p-1} \vec{u}_i \vec{u}_i^T\right) y$$

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


y
-1
1
x
Kernel Trick
-1
1