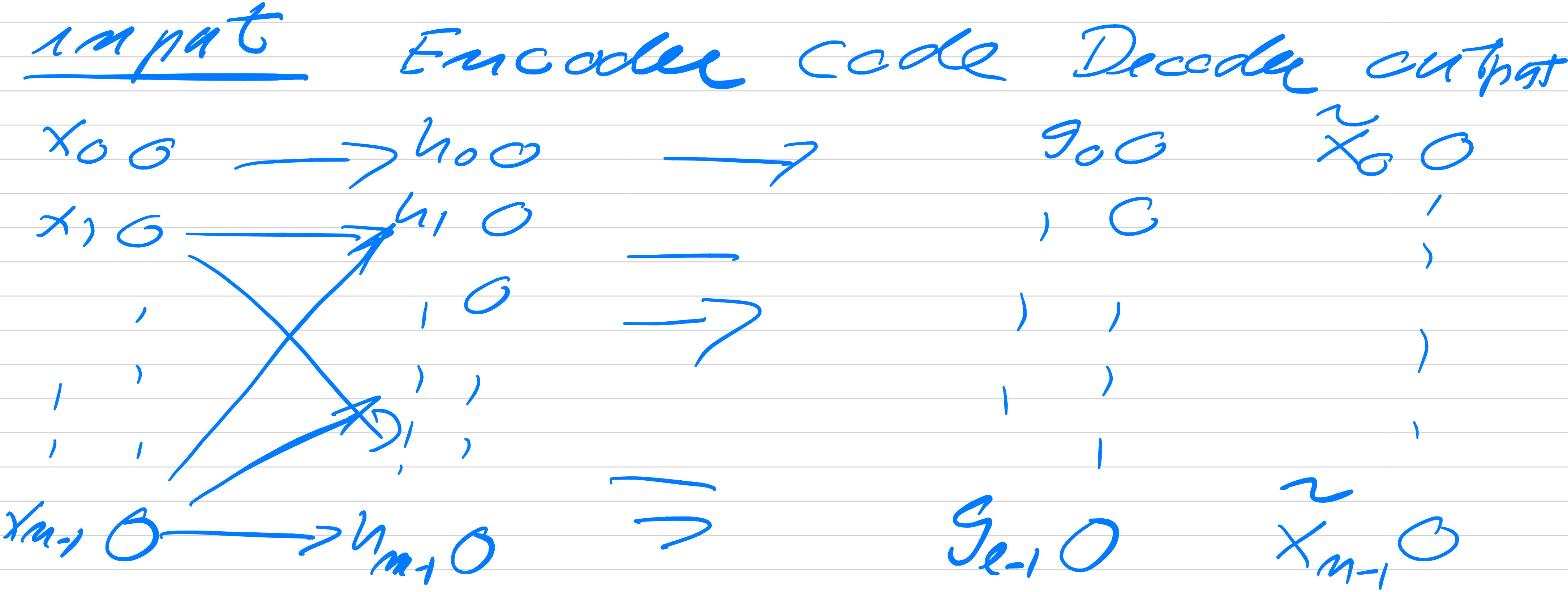
Lecture FYS5429, March 12, 2024

FYS 5429/9429 MARCH 12, 2024



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

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

Define a metric

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

A simple linear dependence

$$\arg\min_{W,V} \frac{1}{n}\sum_{i=0}^{n-1}(x_i - \tilde{x}_i)^2$$

$$h = Wx \qquad \tilde{x} = V \cdot W \cdot x$$

$$g = V \cdot h$$

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

Dimensionality Reduction

1) Example: ordinary least square

$$\tilde{X} = X\Theta$$

X feature matrix

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

# of samples

features

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

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

$$\tilde{x} = X\hat{\theta} = \underbrace{X \frac{1}{\underbrace{X^TX}_{\mathbb{R}^{p\times p}}} X^T}_{A \in \mathbb{R}^{n\times n}} x$$

$$A^2 = A \in \mathbb{R}^{n \times n}$$

$$\tilde{x} = Ax$$

NO - dim reduction

$$\arg\min_{\Theta} \frac{1}{n} \|(I - A^2)x\|_2^2$$

$$= \frac{1}{n} \|(I - A)x\|_2^2$$

autoencoder

$$\arg\min_{W,V} \frac{1}{n} \left(\|(I - \overbrace{VW}^{A})x\|_2^2\right)$$

2) SVD

$$\hat{\epsilon} = \frac{1}{\cancel{X^T X}} X^T x$$

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

$$n \gg p$$

$$X^T X = U D V^T$$

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

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

$$u_i^T u_j = \delta_{ij} \qquad v_i^T v_j = \delta_{ij}$$

$$\mathcal{U} = \begin{bmatrix} | & & | \\ u_0 & \cdots & u_{n-1} \\ | & & | \end{bmatrix}$$

$$D = \begin{bmatrix} \sigma_0 & & & \\ & \sigma_1 & 0 & \\ & 0 & \ddots & \\ & & & \sigma_{p-1} \\ & & & & 0 \\ & & & & & \ddots \end{bmatrix}$$

$$\boxed{\sigma_0 > \sigma_1 > \cdots > \sigma_{p-1} > 0}$$

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

$$D^T D = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \qquad D D^T = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$X^T X = U D V^T$$

$$\tilde{x} = X \frac{1}{X^T X} X^T x =$$

$$= U D V^T \left[\frac{1}{\underbrace{V D^T \underbrace{U^T U}_{I} D V^T}_{p \times p}} \right] V D^T U^T x$$

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

linear algebra gem

if A & B are square matrices and invertible

$$\frac{1}{AB} = B^{-1}A^{-1} \qquad \begin{aligned} V^TV &= \mathbb{1} \\ V &= (V^T)^{-1} \end{aligned}$$

V is a square matrix and invertible

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

$$\frac{1}{VD^TDV^T}$$

$$\tilde{x} = UD\underbrace{V^TV}_{\mathbb{1}}\frac{1}{D^TD}\underbrace{V^TV}_{\mathbb{1}}D^TU^Tx$$

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

with SVD we have a projection with p-degrees of freedom. We want an $m \le p$. Can we find such a value? Yes, PCA

Dim-Reduction and PCA

$$\| x - VWx \|_2^2 \leq \varepsilon$$

PCA tells us that we can define $m \leq p$. How do we define m?

Basic elements of PCA analysis

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

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

$$\mathrm{cov}[X_i, X_j] = \frac{1}{n} \sum_k (x_{ki} - \bar{x}_i) \times (x_{kj} - \bar{x}_j)$$

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

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

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

x_{ki}

all data centered

Example

$$
X = \begin{array}{cc} X_0 & X_1 \\ \left[\begin{array}{c} x_{00} \\ x_{10} \\ x_{20} \end{array}\right. & \left.\begin{array}{c} x_{01} \\ x_{11} \\ x_{21} \end{array}\right] \end{array}
$$

$$
\operatorname{cov}[X_0, X_1] = \frac{1}{3}\left(x_{00}x_{01} + x_{10}x_{11} + x_{20}x_{21}\right)
$$

$n-1$

$$
= X_0^T X_1 \cdot \frac{1}{n}
$$

covariance of X $(X^T X)$

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

$p = 2$

$$\operatorname{Cov}[X] = \frac{1}{n} \begin{bmatrix} x_0^T x_0 & x_0^T x_1 \\ x_1^T x_0 & x_1^T x_1 \end{bmatrix}$$

$$\sigma_i^2 = \frac{1}{n} \sum_k x_{ki}^2$$

$$\mathrm{Cov}[x] = \begin{bmatrix} \sigma_0^2 & \mathrm{cov}[x_0, x_1] \\ \mathrm{cov}[x_1, x_0] & \sigma_1^2 \end{bmatrix}$$

$\mathrm{Corr}[x]$ which has elements

$$\frac{\mathrm{cov}[x_i, x_j]}{\sqrt{\mathrm{var}[x_i]\,\mathrm{var}[x_j]}}$$

$$\mathrm{Corr}[x] = \begin{bmatrix} 1 & & \mathrm{corr}[x_i, x_j] \\ & 1 & \\ & & \ddots \end{bmatrix}$$

Link with SVD

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

$X^T X$ with SVD $\quad X = UDV^T$

$$\mathrm{cov}[X] \quad \frac{1}{n}\left[V D^T \underbrace{U^T U}_{I} D V^T \right]$$

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

$$\frac{1}{n}\left(X^T X\right) V = \frac{1}{n}\left(V D^2 V^T\right) V$$

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

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

$$\frac{1}{n}\left(X^T X\right) v_i = v_i \, \sigma_i^2$$

eigenvalues of $\mathrm{cov}[X]$

are the singular values σ_i^2 with eigenvectors v_i.

Total variance is

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

The PCA constructs a set of orthogonal vectors

$$S = \begin{bmatrix} | & | & & | \\ s_0 & s_1 & \cdots & s_{p-1} \\ | & | & & | \end{bmatrix}$$

$$\bar{S}^T \operatorname{cov}[x]\, S = \operatorname{cov}[\bar{y}]$$

$$\left(\bar{S}^T A S = D \right)$$

which is diagonal

so that

$$\sum_{i=0}^{m-1} \sigma_i^2 \leq$$ Truncated total variance.

Define a reconstruction error

$$J(x) = \frac{1}{n} \| (x - s_0 s_0^T x) \|_2^2$$

$$\mathcal{L} = s_0^T \operatorname{cov}[x] s_0 + \lambda_0 (1 - s_0^T s_0)$$

optimize wrt to s_0 and λ_0

$$\operatorname{cov}[x] s_0 = \lambda_0 s_0$$

multiply with s_0^T

$$s_0^T \operatorname{cov}[x] s_0 = \lambda_0 = \sigma_0^2 / n$$

next: s_1

$$\mathcal{L} = s_1^T \operatorname{cov}[x]\, s_1 + \lambda_1 (1 - s_1^T s_1)$$

$$+ \gamma\, s_1^T s_0$$

Take derivatives wrt λ_1, γ and $\underline{s_1}$

wrt s_1

$$\operatorname{cov}[x]\, s_1 + \gamma/2\, s_0 = \lambda_1 s_1$$

$$s_1^T \operatorname{cov}[x]\, s_1 + \gamma/2\, s_1^T s_0 = \lambda_1$$

$$s_1^T \mathrm{cov}[x] \underbrace{s_1}_{v_1} = \lambda_1$$