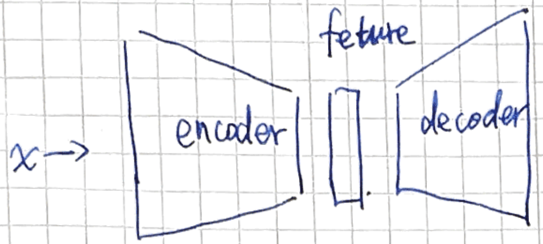
Ex 5-1

a) Autoencoder ~~tries~~ aims to learn an inverse function (feture map) $h = g_e(x, \theta_e)$, an encoder .

But We cannot measure h directly,

① learn feture map.
② Reduce dimentionality



$$L(\theta) = \sum \| x - x' \|^2 .$$

---

b) - ~~Reduce~~ dimentionality Reduction
- Rretraining . Restricted Bolzmamm Machine .
- outlier detection .
- generative model . (GAN) .

---

c) Directly copy the input ..

BRUNNEN

Ex 5-2

a) $K(x,y) = \langle x,y\rangle^2 = (x_1 y_1 + x_2 y_2)^2$

$$= x_1^2 y_1^2 + 2\, x_1 x_2 y_1 y_2 + x_2^2 y_2^2 .$$

$$= \left\langle \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix} \right\rangle .$$

$$= \langle \phi(x), \phi(y) \rangle .$$

2) $K(x,y) = \exp\{-\gamma \|x-y\|^2\}$ for $x,y \in \mathbb{R}$ and $\gamma > 0$.

$$= \exp\{-\gamma (x-y)^2\}$$

$$= \exp\{-\gamma x^2 + 2xy\gamma - \gamma y^2\}.$$

$$= \exp\{-\gamma(x^2+y^2)\} \cdot \exp\{2xy\gamma\}.$$

$$= \exp\{-\gamma(x^2+y^2)\} \cdot \sum_{k=0}^{\infty} \frac{(2xy\gamma)^k}{k!}.$$

$$= \exp\{-\gamma x^2\} \cdot \exp\{-\gamma y^2\} \cdot \left[\sum_{k=0}^{\infty} \left(\sqrt{\frac{2\gamma^k}{k!}} \cdot x^k \cdot \sqrt{\frac{2\gamma^k}{k!}} \cdot y^k\right)\right]$$

$$= \left\langle \exp\{-\gamma x^2\} \begin{pmatrix} \sqrt{\frac{2\gamma}{0!}}\, x^0 \\ \vdots \\ \sqrt{\frac{2\gamma^k}{k!}}\, x^k \\ \vdots \end{pmatrix}, \exp\{-\gamma y^2\} \begin{pmatrix} \vdots \end{pmatrix} \right.$$

$$= \langle \phi(x), \phi(y) \rangle$$

Ex 5-3 Mercel Therem (Sufficient & unnecessary condition)

if Mercel holds then valid

if valid the Merce may not hold

To prove a kernel is a valid kernel:

Method 1. find: $K(x,y) = \phi(x)^T \cdot \phi(y)$.

Method 2. check $K(x,y) = K_1(x,y) \circ K_2(x,y)$

where K_1 K_2 are valid kerels.

$\circ$ is an ~~operation~~ operator.

a) $a K_1(x_i, x_j) = a \phi_0(x_i)^T \phi_0(x_j) = \sqrt{a}\phi_0(x_i)^T \sqrt{a}\phi_0(x_j)$

(Mercer kernel)

$= \phi_0'(x_i)^T \cdot \phi_0'(x_j)$

where $\phi_0'(x) = \sqrt{a}\phi_0(x)$.

b) $K(x_i, x_j) = K_1(x_i, x_j) + K_2(x_i, x_j)$

$= \phi_1(x_i)^T \phi_1(x_j) + \phi_2(x_i)^T \phi_2(x_j)$

$= (\phi_1(x_i), \phi_2(x_i))^T \cdot (\phi_1(x_j), \phi_2(x_j))$

$= \phi'(x_i)^T \phi'(x_j)$ where $\phi'(x) = (\phi_1(x), \phi_2(x))$

c) $K(x_i, x_j) = \sum_{l=1}^{d} w_l K_l(x_i, x_j) \overset{(a)}{=} \sum_{l=1}^{d} \phi(x_i)^T \cdot \phi(x_j)$

$\overset{(b)}{=} \phi'(x_i)^T \cdot \phi'(x_j)$

where $\phi'(x) = \begin{pmatrix} \sqrt{w_1}\phi_1(x) \\ \vdots \\ \sqrt{w_l}\phi_l(x) \end{pmatrix}$

Or by deduction. For $d \in \mathbb{N}^+$,

(1) $w_l K_l(x_i, x_j)$ is valid,

d) $k(x_i, x_j) = k_1(x_i, x_j) \cdot k_2(x_i, x_j)$

$$= \phi_1(x_i)^T \cdot \phi_1(x_j) \cdot \phi_2^T(x_i)\, \phi_2(x_j))$$

$$= \cancel{\phi_1(x_i)^T \phi_2(x_i)^T \cdot \phi_1(x_j)\, \phi_2(x_j))}$$

$$= \left[\sum_{m=1}^{M_1} \phi_1(x_i)_m \phi_1(x_j)_m\right]\left[\sum_{n=1}^{M_2} \phi_2(x_i)_n\, \phi_2(x_j)_n\right]$$

$$= \sum_{m=1}^{M_1} \left[\phi_1(x_i)_m\, \phi_1(x_j)_m \cdot \sum_{n=1}^{M_2} \phi_2(x_i)_n \cdot \phi_2(x_j)_n\right]$$

$$= \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \left[\phi_1(x_i)_m \cdot \phi_2(x_j)_m \cdot \phi_2(x_i)_n \cdot \phi_2(x_j)_n\right]$$

$$= \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \left[(\phi_1(x_i)_m \cdot \phi_2(x_i)_n) \cdot (\phi_1(x_j)_m \cdot \phi_2(x_j)_m\right]$$

$$= \cancel{\sum_{m=1}^{M_1} \phi_1(x_i) \cdot \phi_2(x_i))}$$

$$= \begin{pmatrix} \phi_1(x_i)_1 \cdot \phi_2(x_i)_1 \\ \phi_1(x_i)_1 \cdot \phi_2(x_i)_2 \\ \vdots \\ \phi_1(x_i)_1 \cdot \phi_2(x_i)_2 \\ \phi_1(x_i)_2\ \phi_2(x_i)_1 \\ \vdots \\ \phi_1(x_i)_{M_1} \cdot \phi_2(x_i)_{M_2} \end{pmatrix} \cdot (\phi_1(x_j)_1\ \phi_2(x_j)_1, \dots, \phi_1(x_j)_{M_1}\ \phi_2(x_j)_{M_2}$$

$$= \phi(x_i)\, \phi(x_j)$$

where $\phi(x) = \begin{pmatrix} \phi_1(x)_1 \phi_2(x)_1 \\ \vdots \\ \phi_1(x)_{M_1} \phi_2(x)_{M_2} \end{pmatrix}$

e) $\cancel{k(x_i, x_j) = \phi_1(x_i) \dots (k_1(x_i, x_j))^p}$ (d) $\cancel{= (\phi(x_i)^T \cdot \phi(x_j))^p}$

$=$

c) $k(x_i, x_j) = \sum_{l=1}^{d} w_l k_l (x_i, x_j)$

Proof by deduction:

For $d \in \mathbb{N}^+$, (i) $d = 1$, $k(x_i, x_j) = w_l k_l (x_i, x_j)$

is a valid kernel. (a)

(ii) Assume $d=n$ gives a valid kernel. $\sum_{l=1}^{n} w_l k_l (x_i, x_j)$

Then $d = n+1$

$$k(x_i, x_j) = \sum_{l=1}^{n+1} w_l k_l (x_i, x_j) = \sum_{l=1}^{n} w_l k_l (x_i, x_j) + w_l k_l (x_i, x_j)$$

is a valid kernel. (b).

#

e) $k(x_i, x_j) = (k_1 (x_i, x_j))^p$

Proof by deduction:

For $p \in \mathbb{N}^+$. $p = 1$, $k(x_i, x_j) = k_1 (x_i, x_j)$ is valid.

Assume $p=n$ gives a valid kernel. $k_1(x_i, x_j)^n$

then for $p = n+1$ $k(x_i, x_j) = (k_1(x_i, x_j))^{n+1} = (k_1(x_i, x_j))^n \cdot k_1(x_i, x_j)$

is a valid kernel (d).