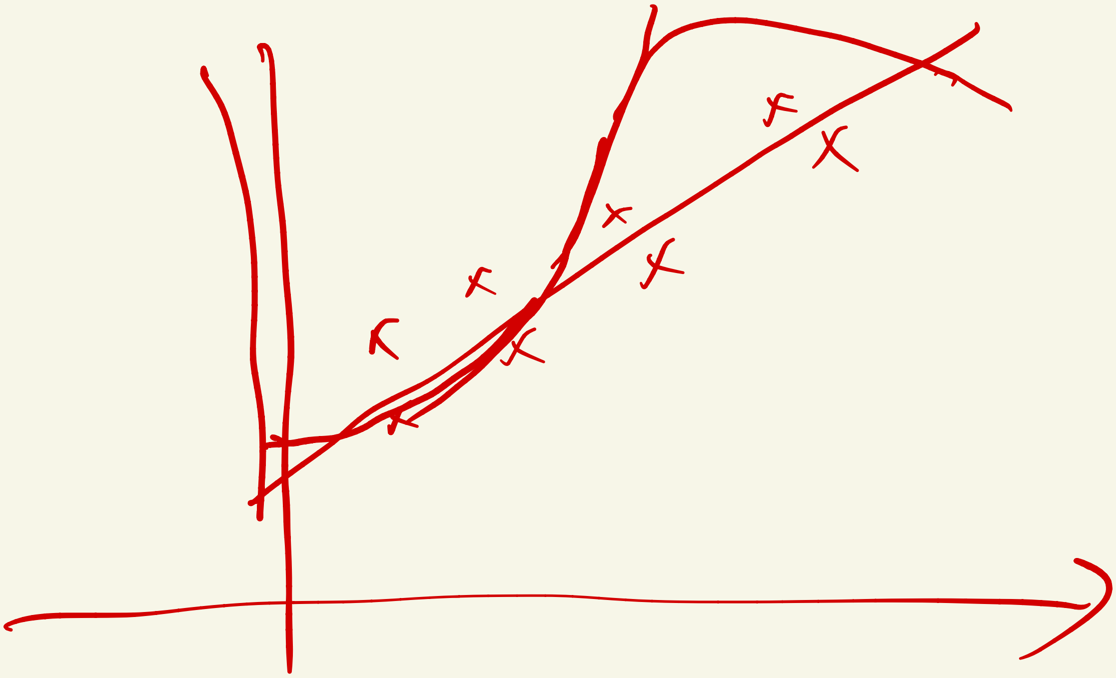
Lecture 4 Kernel Method



d: dimension of input

p: dimension of new feature

$\theta \in R^d$ $\theta \in R^p$

$\phi(\ )$ ← no parameter

representation Learning

ICLR

pixels

x

sofenax

feature mapping

1024

sofemax

$x = [x_1 \quad x_2 \quad x_3]$

$y = \theta^T \underline{\underline{\phi}}(x)$ 3-dim

x^{100}

$y = \theta^T \phi(x)$ p-dim

$\theta_0 = 0$

n: # samples

$$\theta = \sum_{i=1}^{n} \beta_i \phi(x^{(i)}) \qquad \beta_i \in \mathbb{R}$$

assumption:

$$\theta_{old} = \sum_{i=1}^{n} \beta_i \phi(x^{(i)}) \rightarrow$$

gradient descent

$$\theta := \theta_{old} + \alpha \sum_{i=1}^{n} (y^{(i)} - \theta^T \phi(x^{(i)})) \phi(x^{(i)})$$

$$= \sum_{i=1}^{n} \beta_i \phi(x^{(i)}) + \alpha \sum_{i=1}^{n} (y^{(i)} - \theta^T \phi(x^{(i)})) \phi(x^{(i)})$$

$$= \sum_{i=1}^{n} \underbrace{(\beta_i + \alpha(y^{(i)} - \theta^T \phi(x^{(i)})))}_{\text{new } \beta_i} \phi(x^{(i)})$$

$$\theta = \sum_{i=1}^{n} \beta_i \phi[x^{(i)}]$$

$$\theta_0 = 0$$

$$\theta'$$

$$\theta' = \theta - \theta_0$$

$$\downarrow$$

$$\theta = \theta' + \theta_0$$

$$\theta = \theta' + \theta_0$$

$$\beta_i := \beta_i + \alpha \left(y^{(i)} - \sum_{j=1}^{n} \beta_j \, \phi(x^{(j)})^T \phi(x^{(i)}) \right)$$

scalar $\mathbb{R}$

unrelated to β_i

$\theta^T \phi(x^{(i)})$

p

mp multiplication

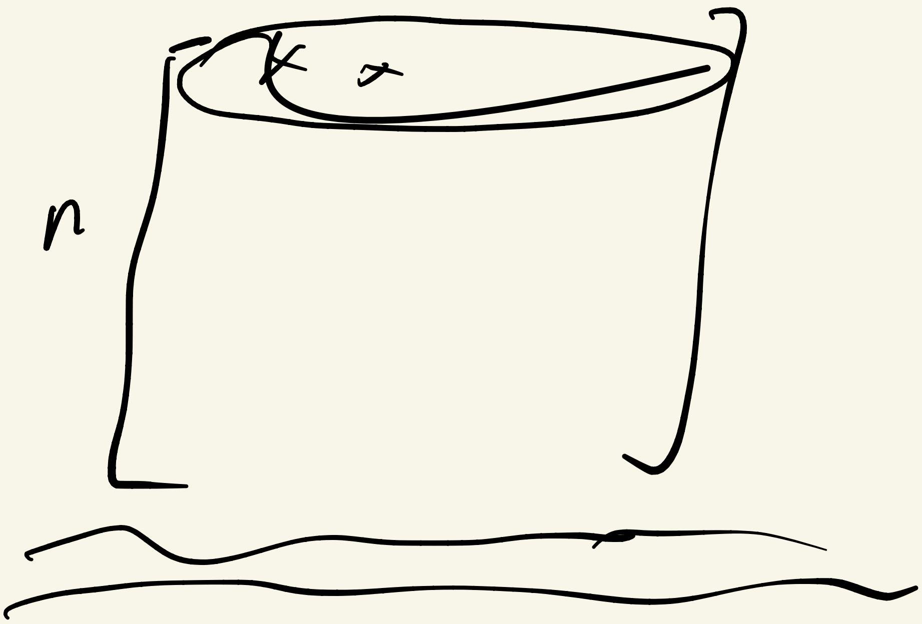
m: # gradient steps

$i, j \in \{1, \cdots n\}$ n: # data samples

100 $\qquad 100^2$

n

$K = \; n$



$K_{ij} = \langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle$

$$\theta = \sum_{i} \beta_i \phi(x^{(i)})$$

$k(x,z) = \phi(x)^T \phi(z)$

$= (x^T z)^2$ $O(3)$

$\phi(x)$?

$x \in R^3$ $d=3$

$\phi(x) \in R^9$ d^2