Kernel Regression

Monday March 8

1. We will do kernel methods first.
2. Gradiance 4 due date extended by 2 days,
3. PA2 will be released on friday Mar 12

Ridge Regression

$$W = (\lambda I_D + X^T X)^{-1} X^T y$$

$$y^* = W^T x^*$$

$$= ((\lambda I_D + X^T X)^{-1} X^T y)^T x^*$$

$$\boxed{y^* = y^T (\lambda I_N + X X^T)^{-1} X x^*}$$

$X \to N \times D$

$X^T \to D \times N$

$$XX^T = N \times N$$

$$XX^T$$

$$x^* \quad \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$$

$$X \quad (4 \times 2)$$

$$\begin{array}{c c} x_1^T & \\ \cancel{x} & \end{array} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \qquad \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} \\ x_{12} & x_{22} & x_{32} & x_{42} \end{bmatrix} X^T$$

$$\begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & \\ \vdots & \ddots & & \\ \vdots & & \ddots & \\ \vdots & & & \ddots \end{bmatrix}$$

$$\boxed{x_1^T x_1 \equiv \langle x_1, x_2 \rangle = \sum_{j=1}^{D} x_{ij}^2}$$

$$Xx^*$$

$$2 \qquad 1$$

$$[\quad] \qquad [\quad] \qquad [\langle x_1, x_2 \rangle]$$

$$\begin{array}{c} \\ \cancel{X} \\ 4 \end{array} \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{array}{c} 2 \\ x^* \end{array} = \begin{bmatrix} \langle x_2, x_* \rangle \\ \langle x_3, x_* \rangle \\ \langle x_4, x_* \rangle \end{bmatrix}$$

$\phi_1(x_1) \longrightarrow$ scalar

$\underset{N \times P}{\underline{\Phi}} \longrightarrow$ is our new data matrix

$$\boxed{y^* = y^T \left(\lambda I_N + \underline{\Phi \Phi^T}\right)^{-1} \underline{\Phi\, \phi(x^*)}}$$

Kernel ~~Trick~~ Method

Dot-product $\langle x_i, x_j \rangle \equiv x_i^T x_j$

is a function of (x_i, x_j)

What if we replace $\langle x_i, x_j \rangle$ with a function $k(x_i, x_j)$

Let K be a $N \times N$ matrix

such that $K[i][j] = k(x_i, x_j)$

and let $k(X, x^*)$ be a $N \times 1$ matrix

such that $\underline{k(\mathbf{X}, x^*)[i]} = k(\mathbf{x}_i, \mathbf{x}_*)$

$$y^* = y^T (\lambda I_N + K)^{-1} k(X, x^*)$$

Can I use any function as a kernel function?

let $D = 2$

$\mathbf{x}_i, \mathbf{x}_j$

$$\boxed{k(x_i, x_j) = x_{i1}^2 x_{j1}^4 - 2 \log\left(\frac{x_{i1} x_{j2}}{4}\right)}$$

No

<u>Direct design</u>

Just come up with a $k(x_i, x_j)$ without doing the basis fn. expansion.

For a $k(\,)$ to be a valid Kernel function:

The K matrix should follow:
$(N \times N)$
$\Downarrow$
Gram matrix
or
kernel matrix

- Symmetric
- Positive semi-definite (p.s.d)

$$\boxed{\begin{gathered}A \text{ is p.s.d if} \\ x^T A x \geq 0 \quad \cancel{\forall x \neq 0}\end{gathered}}$$

If $k(\,)$ is a valid kernel function:

Then there exists a basis function expansion of x_i and x_j such that $k(x_i, x_j) = \Phi(x_i)^T \phi(x_j)$

Kernel Trick

<u>RBF</u> :

$$K(x_i, x_j) = \exp\left[-\frac{1}{2\gamma^2}\|x_i - x_j\|^2\right]$$

$\boxed{\gamma > 0}$

<u>Cosine</u> :

$$\frac{x_i^T x_j}{\|x_i\| \, \|x_j\|}$$

<u>Wed March 10</u>

- Gradiance 5 Out tonight
- PA2 Out on Friday

Kernel function

Kernel Trick

Kernel / Gram matrix

Positive-semi Definite Matrices

$$\boxed{\underline{\cancel{x}^T A x} \geq 0} \quad A \to \text{p.s.d}$$

strategy!

$\cancel{x}$ Assume $x \in R^D$

$\underline{\Phi(x)} \in R^P$

$\underline{W^T \Phi(x)}$

$\underline{W}^T x$

$\underline{k(x_i, x_j)} =$

$\underline{\Phi(x_i)^T \Phi(x_j)}$

$$\boxed{x_i^T \cancel{x}_j = k(x_i, x_j)}$$

<u>Kernel Method</u> – apply implicit basis fn. expansion.

$$k(x_i, x_j) = \exp\left[-\frac{1}{2\gamma^2} \|x_i - x_j\|^2\right]$$

assume $\gamma = 1$ and $\cancel{x}_i, \cancel{x}_j \in R$

$$k(x_i, x_j) = \exp\left[-\frac{1}{2} (x_i - x_j)^2\right]$$

$$= \exp\left[-\frac{1}{2}\left(x_i^2 + x_j^2 - 2x_i x_j\right)\right]$$

$$= \exp(-x_i^2)\ \exp(-x_j^2)\ \exp(2x_i x_j)$$

$$\boxed{e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}}$$