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

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

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

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

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

$X \rightarrow N \times D$

$X^T \rightarrow D \times N$

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

$$XX^T$$

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

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

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

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

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

$$X x^*$$

$$\begin{bmatrix} & 2 & \end{bmatrix} \qquad \begin{bmatrix} 1 \end{bmatrix} \qquad \begin{bmatrix} \langle x_1, x_2 \rangle \end{bmatrix}$$

$$\cancel{X}\;\begin{bmatrix} & \\ & \\ & \end{bmatrix}_{4} \begin{bmatrix} \\ \end{bmatrix}_{x^*} = \begin{bmatrix} \langle x_2, x_* \rangle \\ \langle x_3, x_* \rangle \\ \langle x_4, x_* \rangle \end{bmatrix}$$

$\phi_1(x_1) \longrightarrow$ scalar

$\underset{N \times P}{\Phi} \longrightarrow$ 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 $\underline{k(X, x^*)}[i] = k(x_i, x_*)$

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

Can I use any function as a kernel function?

let $D = 2$

x_i , 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)

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

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

RBF:

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

$$\boxed{\gamma > 0}$$

Cosine:

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