KERNELS (CH. 6)

$\underline{x} \longrightarrow \phi(\underline{x})$ FEATURE SPACE

KERNEL $k(\underline{x}, \underline{x}') = \phi(\underline{x})^T \phi(\underline{x}') = k(\underline{x}', \underline{x})$

$\left(\text{IF } \phi(\underline{x}) = \underline{x} \quad k(\underline{x}, \underline{x}') = \underline{x}^T \underline{x}' \right)$ $\qquad f(\underline{x}^T \underline{x}')$

DUAL REPRESENTATION (CH. 6.1)

(LIN. REGRESSION)

$$J(\underline{w}) = \frac{1}{2} \sum_{n=1}^{N} \{ \underline{w}^T \phi(\underline{x}_n) - t_n \}^2 + \frac{\lambda}{2} \underline{w}^T \underline{w}$$

$\frac{\partial}{\partial \underline{w}} = \phi \longrightarrow \underline{w} =$ LIN. COMBINATION $\{ \phi(\underline{x}_n) \}$

$$\underline{w} = -\frac{1}{\lambda} \sum_{n=1}^{N} \{ \underline{w}^T \phi(\underline{x}_n) - t_n \} \phi(\underline{x}_n) = \sum_{n=1}^{N} a_n \phi(\underline{x}_n) = \underline{\underline{\Phi}}^T \underline{a}$$

$$\underline{\underline{\Phi}} = \begin{bmatrix} \phi(\underline{x}_1)^T \\ \vdots \\ \phi(\underline{x}_N)^T \end{bmatrix} \quad \text{DESIGN MATRIX}$$

$$a_n = -\frac{1}{\lambda} \{ \underline{w}^T \phi(\underline{x}_n) - t_n \}$$

$$J(\underline{w}) = J(\underline{\underline{\Phi}}^T \underline{a}) = J(\underline{a}) =$$

$$\frac{1}{2} \underline{a}^T \underline{\underline{\Phi}} \underline{\underline{\Phi}}^T \underline{\underline{\Phi}} \underline{\underline{\Phi}}^T \underline{a} - \underline{a}^T \underline{\underline{\Phi}} \underline{\underline{\Phi}}^T \underline{t} + \frac{1}{2} \underline{t}^T \underline{t} + \frac{\lambda}{2} \underline{a}^T \underline{\underline{\Phi}} \underline{\underline{\Phi}}^T \underline{a}$$

LET $\underline{\underline{\Phi}} \underline{\underline{\Phi}}^T = \underline{\underline{K}}$ GRAM-MATRIX $\quad (\underline{\underline{K}})_{nm} = \phi(\underline{x}_n)^T \phi(\underline{x}_m) = k(\underline{x}_n, \underline{x}_m)$

$\frac{\partial}{\partial \underline{a}} = \phi \qquad \underline{a} = (\underline{\underline{K}} + \lambda \underline{\underline{I}}_N)^{-1} \underline{t}$

(LIN. REGRESSION) MODEL: (KERNELS INSTEAD OF DATA OR FEATURES)

$$y(\underline{x}) = \underline{w}^T \phi(\underline{x}) = \underline{a}^T \underline{\underline{\Phi}} \phi(\underline{x}) = \underline{k}(\underline{x})^T (\underline{\underline{K}} + \lambda \underline{\underline{I}}_N)^{-1} \underline{t}$$

$$[\cdot]_n = k(\underline{x}_n, \underline{x})$$

GAIN: WORKING WITH KERNEL INSTEAD OF $\underline{x}$, OR DIRECT $\phi(\underline{x})$

CONSTRUCTING KERNELS (CH. 6.2) (EXAMPLE)

$k_0(\underline{x}, \underline{z}) = \underline{x}^T \underline{z} \qquad k(\underline{x}, \underline{z}) = (\underline{x}^T \underline{z})^2 = (x_1 z_1 + x_2 z_2)^2$

$$= x_1^2 z_1^2 + 2 x_1 z_1 x_2 z_2 + x_2^2 z_2^2 =$$

$$= (x_1^2, \sqrt{2} x_1 x_2, x_2^2)(z_1^2, \sqrt{2} z_1 z_2, z_2^2) = \phi(\underline{x})^T \phi(\underline{z})$$