

3.1.1

$$P(w) = \frac{1}{2} \|Xw - y\|^2 + \frac{1}{2} \|w\|^2$$

$$f(x) = \frac{1}{2} \|x - y\|^2 \quad g(w) = \frac{1}{2} \|w\|^2$$

$$g^*(z) = \sup_x \left\{ x^T z - \frac{1}{2} \|x\|^2 \right\}$$

$$\frac{\partial}{\partial x} = 0 \Rightarrow z - \lambda x = 0 \Rightarrow x = \frac{z}{\lambda}$$

$$g^*(z) = \frac{1}{\lambda} z^T z - \frac{1}{2} \left\| \frac{z}{\lambda} \right\|^2$$

$$= \frac{1}{\lambda} \|z\|^2 - \frac{1}{2\lambda} \|z\|^2$$

$$= \frac{1}{2\lambda} \|z\|^2$$

$$D(z) = \underline{\hspace{2cm}} \quad - \frac{1}{2\lambda} \|X^T z\|^2$$

3.1.2

If  $f(x) = \|x\|$ ,

then  $f^*(y) = \begin{cases} 0 & \|y\| \leq 1 \\ \infty & \text{otherwise} \end{cases}$

or the dual norm to  $f(x)$  is general

If  $h(x) = \alpha \cdot f(x)$

then  $h^*(y) = \alpha \cdot f^*\left(\frac{1}{\alpha} y\right)$

If  $h(x) = f(x - y)$

then  $h^*(z) = f^*(z) + y^T z$

### 3.1.3

$$\sup_x \left\{ z^T x - \sum_i f(x_i) \right\} = \sum_i \sup_{x_i} \left\{ z_i x_i - f(x_i) \right\}$$

- Watch for constraints on  $z_i$

### 3.2

while  $\rho - 0 > \epsilon$  e.g.  $10^{-3}$

- Choose random coordinate  $i$

- Update  $z_i$

$$D(z) = e^T z - \frac{1}{2\lambda} z^T G z$$

$$\frac{\partial D(z)}{\partial z_i} = 1 - \frac{1}{\lambda} z^T G^i$$

$$= 1 - \frac{1}{\lambda} \sum_j z_j G_{j,i}^i$$

- Set to 0, solve for  $z_i$

### 3.3.2

$$\text{rbf}(X, X) = K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} K_1 & K_2 \\ \text{rbf}(X, X^{1:m}) & \text{rbf}(X, X^{m+1:n}) \end{bmatrix}$$

$\text{rbf}(X^{m+1:n}, X^{1:m})$

Train

$$z = (K_1^T + K_1 + \lambda \cdot K_{11})^{-1} K_1^T y$$

Test

$$\hat{K} = \text{rbf}(\hat{X}, X^{1:m})$$

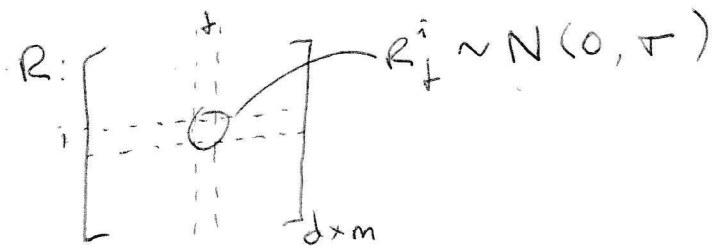
$$\hat{y} = \hat{K} z$$

3.3.3

Train

$$Z = e^{iX} R$$

$$i = \sqrt{-1}$$



$$w = (Z^T Z + \lambda I)^{-1} Z^T y$$

Test

$$\hat{Z} = e^{i\hat{X}} R$$

$$\hat{y} = \hat{Z} \cdot w$$