3.1.1

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

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

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

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

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

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

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

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

3.1.2

If $f(x) = \|x\|_1$

then $f^*(y) = \begin{cases} 0 & \|y\|_\infty \le 1 \\ \infty & \text{otherwise} \end{cases}$ or the dual norm to $f(x)$ in 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 $P - D > \overset{\text{e.g. } 10^{-3}}{\varepsilon}$

- Choose random coordinate i
- Update z_i

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

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

$$= 1 - \frac{1}{\lambda} \sum_j \mathbf{z}_j G^i_j$$

- set to 0, solve for z_i

3.3.2

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

Train

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

Test

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

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

3.3.3

Train

$$Z = e^{iXR}$$

$i : \sqrt{-1}$

$$R : \begin{array}{cc} & t \\ i & \begin{bmatrix} & & \\ & R^i_t & \\ & & \end{bmatrix}_{d \times m} \end{array} \qquad R^i_t \sim N(0, \sigma)$$

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

Test

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

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