

(Q1 a)

$$J(\omega, b) = \frac{1}{n} \sum (y_i - \omega^T x_i + b)^2 + \lambda \|\omega\|^2$$

~~$$J(\theta) = \frac{1}{n} \sum (y_i - \theta^T x_i)$$~~

~~$$J(\theta) = (Y - X\theta)^T (Y - X\theta) + \lambda n \|\theta\|^2$$~~

$$\theta = [b \quad \omega_1 \quad \dots \quad \omega_n]^T \quad A = \begin{bmatrix} 1 & \dots & 1 \\ x_1^T & \dots & x_n^T \end{bmatrix}$$

~~$$\frac{dJ(\theta)}{d\theta} = -2X^T Y + 2(X^T X)\theta + 2\lambda n A^T A \theta = 0$$~~

$$\Rightarrow \theta (X^T X + \lambda n A^T A) = X^T Y$$

$$\therefore \theta = (X^T X + \lambda n A^T A)^{-1} X^T Y$$

~~$$\& y = \theta^T x$$~~

(Q1 a) No, they won't be equivalent. If we center the data in input space & then apply KRR without offset, the form of kernel will be  $\langle \phi(x_i - \bar{x}), \phi(v - \bar{x}) \rangle$  whereas in case of KRR without offset, kernel will be  $\langle (\phi(x_i) - \bar{\phi}(x)), (\phi(v) - \bar{\phi}(x)) \rangle$ . And these two expressions are different.  
Hence they will be different.

$$\begin{aligned} b) \quad b &= \bar{y} - \omega^T \bar{x} \\ &= \bar{y} - \perp_{\mu} \bar{y}^T [x - \tilde{K}(\tilde{K} + \mu I)^{-1} x] \bar{x} \\ &= \bar{y} - \perp_{\mu} \bar{y}^T [I - \tilde{K}(\tilde{K} + \mu I)^{-1}] \tilde{K}(x) \end{aligned}$$

where  $K(x) = \begin{bmatrix} \langle \bar{x}_1, \bar{x}_1 \rangle \\ \langle \bar{x}_1, \bar{x}_2 \rangle \\ \vdots \\ \langle \bar{x}_1, \bar{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \langle x_1 - \bar{x}, \bar{x} \rangle \\ \langle x_2 - \bar{x}, \bar{x} \rangle \\ \vdots \\ \langle x_n - \bar{x}, \bar{x} \rangle \end{bmatrix}$

c)  $\sim K$

The ij element of  $\tilde{K}'$  is

$$\begin{aligned} \langle x_i - \bar{x}, x_j' - \bar{x}^* \rangle &= \langle x_i, x_j' \rangle - \frac{1}{n} \sum_{i=1}^n \langle x_i, x_i \rangle \\ &\quad - \frac{1}{n} \sum_{i=1}^n \langle x_i, x_j' \rangle + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \langle x_i, x_j \rangle \\ &= K' - K O_{n \times n} - O_{n \times n} K' + O_{n \times n} K O_{n \times n} \end{aligned}$$

where  $K$  is the train-train kernel matrix & each element of  $O$  is  $Y_n$  & dimensions of  $O$  matrix are different from those given specified wherever they are mentioned.

Also, for feature map i.e.  $\phi(x)$  just replace  $x_i$  with  $\phi(x_i)$

For predictions.

$$\hat{f}(x) = \bar{y} + \tilde{X}' \tilde{w}$$

where  $\bar{Y}$  is a vector  $\begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix}$   $\bar{y}$  = mean of  $y$  for train samples

$$2 \tilde{X}' = \begin{bmatrix} x_1' - \bar{x} \\ x_2' - \bar{x} \\ \vdots \\ x_m' - \bar{x} \end{bmatrix} \quad \bar{x} = \text{mean vector of } x_i \text{ for training samples.}$$

$$\begin{aligned} \therefore \hat{f}(x) &= \bar{Y} + \tilde{X}' \left[ \frac{1}{m} X^T \left[ I - (\tilde{K} + m\tilde{I})^{-1} \tilde{K} \right] \tilde{y} \right] \\ &= \bar{Y} + \frac{1}{m} (\tilde{X}' \tilde{X}) \left[ I - (\tilde{K} + m\tilde{I})^{-1} \tilde{K} \right] \tilde{y} \\ &= \bar{Y} + \frac{1}{m} (\tilde{R}') \left[ I - (\tilde{K} + m\tilde{I})^{-1} \tilde{K} \right] \tilde{y} \end{aligned}$$

Typo

d)  $MSE_{Train} = 10.3465$

$$MSE_{Test} = 49.2640$$

$$b = 15.9750$$

e)  $MSE_{Train} = 16.4636$

$$MSE_{Test} = 161.0847.$$

The KRR without offset is performing worse than KRR with offset. This also seems logical because KRR without offset doesn't provide the best possible fit as it is constrained to pass through origin in the feature space.

$\omega, \xi_i^+, \xi_i^-$  $J_{\text{SVM}} =$ 

$$(b) \min_{\omega, b, \xi^+, \xi^-} \frac{1}{2} \|\omega\|^2 + \frac{C}{n} \sum (\xi_i^+ + \xi_i^-)$$

$$\text{s.t. } \begin{aligned} & y_i - \omega^\top x_i - b \leq \epsilon + \xi_i^+ \quad \forall i \quad (1) \\ & \omega^\top x_i + b - y_i \leq \epsilon + \xi_i^- \quad \forall i \quad (2) \\ & \xi_i^+ \geq 0 \quad \forall i \\ & \xi_i^- \geq 0 \quad \forall i \end{aligned}$$

From (1) &amp; (2)

$$\begin{aligned} \xi_i^+ &\geq y_i - \omega^\top x_i - b - \epsilon \quad \& \quad \xi_i^+ \geq 0 \\ \xi_i^- &\geq \omega^\top x_i + b - y_i - \epsilon \quad \& \quad \xi_i^- \geq 0 \end{aligned}$$

Let

$$(y_i - \omega^\top x_i - b) \leq -\epsilon,$$

$$\therefore \xi_i^+ \geq 0 \geq y_i - \omega^\top x_i - b - \epsilon$$

$$\therefore \xi_i^+ = 0$$

- (3)

$$\xi_i^- \geq \omega^\top x_i + b - y_i - \epsilon \geq -\epsilon + \epsilon$$

$$\xi_i^- \geq \omega^\top x_i + b - y_i - \epsilon \geq 0$$

$$\xi_i^- = \omega^\top x_i + b - y_i - \epsilon$$

- (4)

or  ~~$\xi_i^- = 0$~~ 

$$\text{Let } (y_i - \omega^\top x_i - b) \geq \epsilon$$

$$\xi_i^+ \geq y_i - \omega^\top x_i - b - \epsilon \geq \epsilon - \epsilon \geq 0$$

$$\therefore \xi_i^+ = y_i - \omega^\top x_i - b - \epsilon$$

- (5)

$$\& \xi_i^- \geq \omega^\top x_i + b - y_i - \epsilon$$

$$\geq -\epsilon - \epsilon$$

$$\geq 0 \geq -2\epsilon$$

$$\therefore \xi_i^- \geq 0$$

- (6)

For ③ & ④

$$\text{If } y_i - \omega^T x_i - b \leq -\epsilon$$

$$\left. \begin{array}{l} \xi_i^+ = 0 \\ \xi_i^- = -y_i + \omega^T x_i + b + \epsilon \\ \quad + \omega^T x_i + b - y_i - \epsilon \end{array} \right\} \quad ⑦$$

From ⑤ & ⑥

$$\text{If } y_i - \omega^T x_i - b \geq \epsilon$$

$$\left. \begin{array}{l} \xi_i^+ = y_i - \omega^T x_i - b + \epsilon \\ \xi_i^- = 0 \end{array} \right\} \quad ⑧$$

$$\text{If } -\epsilon \leq y_i - \omega^T x_i - b \leq +\epsilon$$

$$\left. \begin{array}{l} \xi_i^+ \geq y_i - \omega^T x_i - b - \epsilon \\ \geq 0 \geq y_i - \omega^T x_i - b - \epsilon \end{array} \right\}$$

$$\therefore \xi_i^+ = 0$$

$$\left. \begin{array}{l} \xi_i^- \geq \omega^T x_i + b - y_i - \epsilon \\ \geq 0 \geq y_i - \omega^T x_i - b - \epsilon \end{array} \right\}$$

$$\therefore \xi_i^- = 0$$

∴ From ⑦, 8 & ⑨

$$\xi_i^+ + \xi_i^- = y_i - \omega^T x_i - b - \epsilon ; \left\{ \begin{array}{l} y_i - \omega^T x_i - b \geq \epsilon \\ 0 \\ \omega^T x_i + b - y_i - \epsilon \end{array} \right. \begin{array}{l} y_i - \omega^T x_i - b \geq \epsilon \\ -\epsilon \leq y_i - \omega^T x_i - b \leq \epsilon \\ y_i - \omega^T x_i - b \leq -\epsilon \end{array}$$

$$\text{or } \xi_i^+ + \xi_i^- = \max \{ 0, |y_i - (\omega^T x_i + b)| - \epsilon \} \\ = \max (y_i, \omega^T x_i + b) - \epsilon$$

$$\therefore J(\omega, b, \xi^-, \xi^+) = \min \frac{1}{2} \|\omega\|^2 + \frac{c}{n} \sum (\xi_i^+ + \xi_i^-)$$

$$= \min \frac{1}{2} \|\omega\|_2^2 + \frac{c}{n} \sum \text{loss}(y_i, \omega^\top x_i + b)$$

$$= \min \frac{1}{n} \sum \text{loss}(y_i, \omega^\top x_i + b) + \frac{1}{2c} \|\omega\|_2^2$$

$\therefore$  Let  $\gamma = \frac{1}{2c}$

$$= \min \frac{1}{n} \sum \text{loss}(y_i, \omega^\top x_i + b) + \gamma \|\omega\|_2^2$$

$$\begin{aligned} b \cdot L(\omega, b, \xi^-, \xi^+, \alpha, \beta, \gamma, \delta) &= \frac{1}{2} \|\omega\|_2^2 + \frac{c}{n} \sum (\xi_i^- + \xi_i^+) \\ &\quad + \sum \alpha_i (y_i - \omega^\top x_i - b - \epsilon - \xi_i^+) \\ &\quad + \sum \beta_i (\omega^\top x_i + b - y_i - \epsilon - \xi_i^-) \\ &\quad - \sum \gamma_i (\xi_i^+) + \sum \delta_i (\xi_i^-) \end{aligned}$$

The dual problem is

$$\max_{\alpha \geq 0, \beta \geq 0, \gamma \geq 0, \delta \geq 0} L_D(\alpha, \beta, \gamma, \delta)$$

where

$$L_D(\alpha, \beta, \gamma, \delta) = \min_{\omega, b, \xi^-, \xi^+} L(\omega, b, \xi^-, \xi^+, \alpha, \beta, \gamma, \delta)$$

$$\frac{\partial L}{\partial \omega} = \omega + \sum \alpha_i (-x_i) + \sum \beta_i x_i = 0$$

$$\therefore \omega = \sum \alpha_i x_i - \sum \beta_i x_i \quad (1)$$

$$\frac{\partial L}{\partial b} = \sum (-\alpha_i) + \sum \beta_i = 0 \quad (2)$$

$$\frac{\partial L}{\partial \xi_i^-} = \frac{c}{n} - \beta_i - \gamma_i = 0 \quad \forall i \quad (3)$$

$$\frac{\partial L}{\partial \xi_i^+} = \frac{c}{n} - \alpha_i - \gamma_i = 0 \quad \forall i \quad (4)$$

Using ③ & ④, we can eliminate  $\xi_i^+$  &  $\xi_i^-$

$$L = \frac{1}{2} \|\omega\|^2 + \sum \alpha_i (y_i - \omega^\top x_i - b) + \sum \beta_i (\omega^\top x_i + b - y_i - \epsilon)$$

$$= \frac{1}{2} \|\omega\|^2 + \sum (y_i - \omega^\top x_i - b) (\alpha_i + \beta_i) + \sum \epsilon (-\alpha_i - \beta_i) \\ (-\alpha_i - \beta_i)$$

$$= \frac{1}{2} \|\omega\|^2 + \sum (y_i - \omega^\top x_i) (\beta_i + \alpha_i) + b \left\{ \sum (\beta_i) - \sum \alpha_i \right\}$$

$$\epsilon \in \left\{ \sum \alpha_i + \sum \beta_i \right\}$$

$$= \frac{1}{2} \|\omega\|^2 + \sum (y_i - \omega^\top x_i) (\beta_i + \alpha_i) - 2\epsilon \sum \alpha_i$$

Note:  ~~$\omega^\top x_i = \left\{ \sum_{j=1}^n (\alpha_j - \beta_j) x_j^\top \right\} x_i$~~

~~$$\|\omega\| = \sum (\alpha_i - \beta_i) x_i^\top x_i (\alpha_i - \beta_i)$$~~

~~$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix}$$~~

~~$$\|\omega\| = (\alpha - \beta)^\top X \cdot X (\alpha - \beta)$$~~

~~$$L = \frac{1}{2} \left( \sum (\alpha_i - \beta_i) x_i \right) \left( \sum (\alpha_j - \beta_j) x_j \right) + \sum y_i (\beta_i + \alpha_i) \\ - \sum (\alpha_i - \beta_i) x_i^\top x_i (\beta_i + \alpha_i) - 2\epsilon \sum \alpha_i$$~~

~~$$= -\frac{1}{2} \sum (\alpha_i - \beta_i) (\alpha_j - \beta_j) x_i^\top x_j + \sum y_i (\alpha_i - \beta_i) \\ - 2\epsilon \sum \alpha_i$$~~

$$\therefore L(\alpha, \beta, \epsilon) = -\frac{1}{2} \sum (\alpha_i - \beta_i) (\alpha_j - \beta_j) \langle x_i, x_j \rangle + \sum y_i (\alpha_i - \beta_i) \\ - 2\epsilon \sum \alpha_i$$

$$\sum (\beta_i - \alpha_i) = 0$$

$$\begin{aligned} c/n - \beta_i - \gamma_i &= 0 \quad \forall i \\ c/n - \alpha_i - \gamma_i &= 0 \quad \forall i \end{aligned}$$

$$\text{or } 0 \leq \alpha_i \leq c/n \quad \& \quad 0 \leq \beta_i \leq c/n \quad \forall i$$

$$\therefore L_D(\alpha, \beta) = -\frac{1}{2} \sum_{i,j} (\alpha_i - \beta_j)(\alpha_i - \beta_j) + \sum_i y_i(\alpha_i - \beta_i)$$

~~$\sum_{i,j} \alpha_i \beta_j$~~

$$-2 \in \sum \alpha_i$$

 $w^t$ 

$$\sum (\beta_i - \alpha_i) = 0$$

$\& \quad 0 \leq \alpha_i \leq c/n \quad \& \quad 0 \leq \beta_i \leq c/n$

Given that the strong duality holds. Therefore using the KKT necessity theorem & its conditions,

(1) From first KKT condition,

$$w^* - \sum \alpha_i x_i + \sum \beta_i x_i = 0$$

$$\text{or } w^* = \sum_{i=1}^n (\alpha_i - \beta_i) x_i$$

(2) From complementary slackness,

$$\alpha_i^* (y_i - w^T x_i - b - e - \xi_i^+) = 0$$

If  $x_i$  satisfies above eq.,

$$\begin{aligned} y_i - w^T x_i - b - e - \xi_i^+ &= 0 \quad \cancel{\textcircled{1}} \\ y_i - (w^T x_i + b) &= e + \xi_i^+ \quad \textcircled{2} \end{aligned}$$

$$\beta_i^* (\gamma_i w^\top x_i + b - y_i - \xi_i^-) = 0$$

If  $x_i$  satisfies above eq. then,

$$w^\top x_i + b - y_i = \epsilon + \xi_i^- \quad (2) \text{ or}$$

$\therefore$  All  $x_i$  which satisfies (1) & (2) will be called support vectors.

$\therefore$  if  $x_i$  is not a SV, then  $\alpha_i^* = 0$  or  
 $w^*$  depends on SVs only

$$\therefore w^* = \sum_{\substack{\text{support} \\ \text{vectors}}} \alpha_i^* \underset{\beta_i^*}{x_i}$$

### (3) From complimentary slackness

$$\gamma_i^* \xi_i^+ = 0$$

$$\& \text{ we know, } \gamma_i^* + \alpha_i^* = c/n$$

ie if  $\alpha_i^* < c/n$ , then  $\gamma_i^* > 0$  &  $\xi_i^+ = 0$   
 i.e. if  $\beta_i^* > c/n$

$$\gamma_i^* \xi_i^- = 0$$

(we know,

$$\gamma_i^* + \beta_i^* = c/n$$

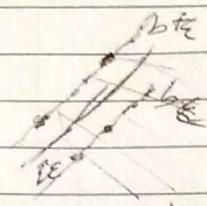
ie if  $\beta_i^* < c/n$ , then  $\gamma_i^* > 0$  &  $\xi_i^- = 0$

$\therefore$  Considering any i s.t.  $0 < \alpha_i^* < c/n$  &  $0 < \beta_i^* < c/n$   
 Then

$$\begin{aligned} w^\top x_i + b - y_i &= \epsilon \Rightarrow b = y_i - w^\top x_i + \epsilon \\ \text{Or } y_i - (w^\top x_i + b) &= \epsilon \Rightarrow b = y_i - w^\top x_i - \epsilon \end{aligned}$$

So, the possible values for  $b^*$  are

$$\begin{aligned} & y_i - \mathbf{w}^\top \mathbf{x}_i + \epsilon \quad \text{or} \quad y_i - \mathbf{w}^\top \mathbf{x}_i - \epsilon \\ & \text{when } \alpha_i < c/n \qquad \qquad \qquad \text{when } \beta_i < c/n \\ & \mathbf{w}^* = \sum_{\text{support vectors}} (\alpha_i^* - \beta_i^*) \mathbf{x}_i \end{aligned}$$



c) The SVR is

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \mathbf{w}^* \mathbf{x} + b^* \\ &= \sum (\alpha_i^* - \beta_i^*) \mathbf{x}_i^\top \mathbf{x} + b^* \\ &= \sum (\alpha_i^* - \beta_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b^* \end{aligned}$$

where  $\mathbf{x}^* = [\alpha_1^*, \dots, \alpha_n^*]^T$  &  $\mathbf{B}^* = [\beta_1^*, \dots, \beta_n^*]^T$   
is the sol. of

$$\max_{\mathbf{w}, b} -\frac{1}{2} \sum (\alpha_i - \beta_i) (\alpha_j - \beta_j) \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum y_i (\alpha_i - \beta_i)$$

$$-\beta \in \sum \alpha_i$$

s.t.

$$\begin{aligned} \sum (\beta_i - \alpha_i) &= 0 \\ &\& 0 \leq \alpha_i \leq K \quad 0 \leq \beta_i \leq c/n \end{aligned}$$

∴  $b^*$  is given by

$$\begin{aligned} & y_i - \mathbf{w}^\top \mathbf{x}_i + \epsilon \quad \text{if} \quad y_i - \mathbf{w}^\top \mathbf{x}_i - \epsilon \\ & \text{where } \alpha_i < c/n \qquad \qquad \qquad \text{when } \beta_i < c/n \end{aligned}$$

d) Only the training examples which are part of support vectors will be considered determine the final predictor because  $\mathbf{w}^*$  won't be depending on parameters

Support vectors are the  $\mathbf{x}_i$ 's which satisfies either of the below conditions  
i.e

$$\mathbf{w}^\top \mathbf{x}_i + b \geq y_i - \epsilon = \xi_i \geq 0$$

DATE

$$\text{or } y_i - (\omega^\top x_i + b) - \epsilon = \tilde{\epsilon}^+ \geq 0$$

ie all the points which satisfies  
<sup>drawing</sup>

$$|y_i - (\omega^\top x_i + b)| \geq \epsilon$$

ie all the points for which predictor error  
is greater than or equal to  $\epsilon$ .