- Lagrangian

-- $L(x,\lambda,\nu) := f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \sum_{i=1}^{n} \nu_j h_j(x)$ where $g_i(x) \le 0 \quad i \in \{1,2,\dots,n\}$ where $h_j(x) = 0 \quad j \in \{1,2,\dots,n\}$

-- Rewrite constraints s.t. they're of the form $g_i(X) \le 0$ for $i = 1, 2, \dots, n$

$$\left[\begin{array}{llll} y_i - w^T x_i - b - \epsilon - \xi_i^+ \le 0 & \forall i & -\xi_i^+ \le 0 & \forall i \\ -y_i + w^T x_i + b - \epsilon - \xi_i^- \le 0 & \forall i & -\xi_i^- \le 0 & \forall i \end{array}\right] \text{REWRITTEN CONSTRAINTS}$$

-- $$L(w,b,\xi_i^+,\xi_i^-,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-)$$
$$+ \sum_{i=1}^{n} \alpha_i^+ (y_i - w^T x_i - b - \epsilon - \xi_i^+) + \sum_{i=1}^{n} \alpha_i^- (-y_i + w^T x_i + b - \epsilon - \xi_i^-)$$
$$+ \sum_{i=1}^{n} \beta_i^+ (-\xi_i^+) + \sum_{i=1}^{n} \beta_i^- (-\xi_i^-)$$

- Dual Function : minimize the Lagrangian wrt primal vars, take ∇ wrt w, b, ξ_i^+, ξ_i^-

$$L_D(\alpha,\beta) = \min_{w,b,\xi_i^+,\xi_i^-} L(w,b,\xi_i^+,\xi_i^-,\alpha,\beta)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i^+ x_i - \sum_{i=1}^{n} \alpha_i^- x_i \overset{set}{=} 0 \longrightarrow \boxed{\cancel{w = 2\sum_{i=1}^{n} \alpha_i x_i}} = \sum_{i=1}^{n} x_i (\alpha_i^+ + \alpha_i^-)$$

$$\frac{\partial L}{\partial b} = -\alpha_i^+ - \alpha_i^- = -2\alpha_i \overset{set}{=} 0 \longrightarrow \boxed{\cancel{\alpha_i = 0}}$$

$$\frac{\partial L}{\partial \xi_i^+} = \frac{C}{n} - \alpha_i^+ - \beta_i^+ \overset{set}{=} 0 \longrightarrow \cancel{\frac{C}{n} - 0 - \beta_i} \Longrightarrow \boxed{\cancel{\beta_i = \frac{C}{n}}}$$

$$\frac{\partial l}{\partial \xi_i^-} = \frac{C}{n} - \alpha_i^- - \beta_i^- \overset{set}{=} 0$$

- Distribute stuff out in Lagrangian, plug in $\underset{\sim}{w}$, other stuff.
- $b\,(\Sigma \text{ stuff}) \longrightarrow$ by KKT, stuff $\overset{should}{=} 0$

$\sum_{i=1}^{n} \xi_i$ (other stuff) $\longrightarrow$ use other KKT so (other stuff) = 0

$\frac{C}{n} - \alpha_i - \beta_i = 0$

- b terms, ξ_i terms
(gather) (gather)

SLDM hw5, 2B

- $L(w, b, \xi_i^+, \xi_i^-, \alpha_i^+, \alpha_i^-, \beta_i^+, \beta_i^-) = \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-)$

(Lagrangian)

$$+ \sum_{i=1}^{n}\alpha_i^+(y_i - w^Tx_i - b - \epsilon - \xi_i^+) + \sum_{i=1}^{n}\alpha_i^-(-y_i + w^Tx_i + b - \epsilon - \xi_i^-)$$

$$+ \sum_{i=1}^{n}\beta_i^+(-\xi_i^+) + \sum_{i=1}^{n}\beta_i^-(-\xi_i^-)$$

- $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i^+ x_i + \sum_{i=1}^{n}\alpha_i^- x_i \overset{set}{=} 0 \longrightarrow \boxed{w = \sum_{i=1}^{n} x_i(\alpha_i^+ - \alpha_i^-)} = \boxed{\sum_{i=1}^{n} x_i\alpha_i^+ - \sum_{i=1}^{n} x_i\alpha_i^-}$

- $\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i^+ + \sum_{i=1}^{n}\alpha_i^- \overset{set}{=} 0 \longrightarrow \boxed{\sum_{i=1}^{n}\alpha_i^+ = \sum_{i=1}^{n}\alpha_i^-}$

- $\frac{\partial L}{\partial \xi_i^+}$: effectively, take $\frac{\partial L}{\partial \xi_i^+}$ of $\left[\frac{C}{n}(\xi_1^+ + \xi_1^- + \xi_2^+ + \xi_2^- + \dots) - (\alpha_1^+\xi_1^+ + \alpha_2^+\xi_2^+) - (\beta_1^+\xi_1^+ + \beta_2^+\xi_2^+ + \dots)\right]$

The partial wrt some ξ_i^+ (some specific i) will be:

$$\frac{C}{n}(0 + \dots + 0 + 1 + 0 + \dots + 0) - (0 + \dots + 0 + 1(\alpha_i^+) + 0 + \dots + 0) - (0 + \dots + 0 + 1(\beta_i^+) + 0 + \dots + 0)$$

$$= \frac{C}{n}(1) - \alpha_i^+ - \beta_i^+ \Longrightarrow \frac{\partial L}{\partial \xi_i^+} = \boxed{\frac{C}{n} - \alpha_i^+ - \beta_i^+ \overset{set}{=} 0}$$

- $\frac{\partial L}{\partial \xi_i^-} = \boxed{\frac{C}{n} - \alpha_i^- - \beta_i^- \overset{set}{=} 0}$

- Distribute stuff out in Lagrangian, plug in w and other stuff.
- Gather "b terms" and "ξ_i terms" such that $b(\Sigma \text{ stuff}) \longrightarrow$ by KKT, stuff should $= 0$

  such that $\sum_{i=1}^{n}\xi_i(\text{other stuff}) \longrightarrow$ use other KKT so that (other stuff) $= 0$

  $\frac{C}{n} - \alpha_i - \beta_i = 0$

- **Expand Lagrangian:** $\frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i^+ + \frac{C}{n}\sum_{i=1}^{n}\xi_i^-$

$$+ \sum_{i=1}^{n}\alpha_i^+ y_i - \sum_{i=1}^{n}\alpha_i^+ w^T x_i - \sum_{i=1}^{n}\alpha_i^+ b - \sum_{i=1}^{n}\alpha_i^+ \epsilon - \sum_{i=1}^{n}\alpha_i^+ \xi_i^+$$

$$- \sum_{i=1}^{n}\alpha_i^- y_i + \sum_{i=1}^{n}\alpha_i^- w^T x_i + \sum_{i=1}^{n}\alpha_i^- b - \sum_{i=1}^{n}\alpha_i^- \epsilon - \sum_{i=1}^{n}\alpha_i^- \xi_i^-$$

$$+ \sum_{i=1}^{n}\beta_i^+(-\xi_i^+) + \sum_{i=1}^{n}\beta_i^-(-\xi_i^-)$$

- **Substitute:** $\frac{C}{n} = \alpha_i^+ + \beta_i^+$ OR $\frac{C}{n} = \alpha_i^- + \beta_i^-$,

Some $\sum_{i=1}^{n}\alpha_i^- = \sum_{i=1}^{n}\alpha_i^+$, giving:

$$\frac{1}{2}\sum_{i=1}^{n} x_i(\alpha_i^+ - \alpha_i^-) + \left[\cancel{\sum_{i=1}^{n}\alpha_i^+\xi_i^+} + \cancel{\sum_{i=1}^{n}\beta_i^+\xi_i^+}\right] + \left[\cancel{\sum_{i=1}^{n}\alpha_i^-\xi_i^-} + \cancel{\sum_{i=1}^{n}\beta_i^-\xi_i^-}\right]$$

$+ \leftarrow$ same as above $\longrightarrow + \dots \; \cancel{-\sum_{i=1}^{n}\alpha_i^+\xi_i^+}$

$- \leftarrow$ same as above $\longrightarrow + \dots \; \cancel{-\sum_{i=1}^{n}\alpha_i^-\xi_i^-}$

$$\cancel{-\sum_{i=1}^{n}\beta_i^+\xi_i^+} \quad \cancel{-\sum_{i=1}^{n}\beta_i^-\xi_i^-}$$

• Now sub in for w: $w = \sum_{i=1}^{n} x_i(\alpha_i^+ - \alpha_i^-)$

$$\frac{1}{2}\left(\sum_{i=1}^{n} x_i(\alpha_i^+ - \alpha_i^-)\right)^T\left(\sum_{i=1}^{n} x_i(\alpha_i^+ - \alpha_i^-)\right)$$

$$+\sum_{i=1}^{n}\alpha_i^+ y_i - \sum_{i=1}^{n}\alpha_i^+(x_i(\alpha_i^+ - \alpha_i^-))^T x_i - \sum_{i=1}^{n}\alpha_i^+ b - \sum_{i=1}^{n}\alpha_i^+\epsilon$$

$$-\sum_{i=1}^{n}\alpha_i^- y_i - \sum_{i=1}^{n}\alpha_i^-(x_i(\alpha_i^+ - \alpha_i^-))^T x_i + \sum_{i=1}^{n}\alpha_i^- b - \sum_{i=1}^{n}\alpha_i^-\epsilon$$

• $$\frac{1}{2}\left[\sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)^2\langle x_i, x_i\rangle\right] = \frac{1}{2}\left[\sum_{i=1}^{n}(\alpha_i^+)^2\|x_i\|^2 - 2\sum_{i=1}^{n}\alpha_i^+\alpha_i^-\|x_i\|^2 + \sum_{i=1}^{n}(\alpha_i^-)^2\|x_i\|^2\right]$$

• We now have: $$\frac{1}{2}\sum_{i=1}^{n}(\alpha_i^+)^2\|x_i\|^2 - \sum_{i=1}^{n}\alpha_i^+\alpha_i^-\|x_i\|^2 + \frac{1}{2}\sum_{i=1}^{n}(\alpha_i^-)^2\|x_i\|^2$$

$$-\sum_{i=1}^{n}(\alpha_i^+)^2\|x_i\|^2 ++ \sum_{i=1}^{n}\alpha_i^+\alpha_i^-\|x_i\|^2 ++ \sum_{i=1}^{n}(\alpha_i^-)^2\|x_i\|^2 - \sum_{i=1}^{n}(\alpha_i^+)(\alpha_i^-)\|x_i\|^2$$

$$+\boxed{\left(\sum_{i=1}^{n}\alpha_i^+ y_i - \sum_{i=1}^{n}\alpha_i^- y_i\right)} + \boxed{\left(-\sum_{i=1}^{n}\alpha_i^+ b + \sum_{i=1}^{n}\alpha_i^- b\right)} + \boxed{\left(-\sum_{i=1}^{n}\alpha_i^+\epsilon - \sum_{i=1}^{n}\alpha_i^-\epsilon\right)}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}(\alpha_i^+)^2\|x_i\|^2 - \sum_{i=1}^{n}(\alpha_i^+)(\alpha_i^-)\|x_i\|^2 + \frac{3}{2}\sum_{i=1}^{n}(\alpha_i^-)^2\|x_i\|^2$$

$$+\boxed{\left(\sum_{i=1}^{n} y_i(\alpha_i^+ - \alpha_i^-)\right)} + \boxed{\left(\sum_{i=1}^{n} b(\alpha_i^- - \alpha_i^+)\right)} + \boxed{\left(\sum_{i=1}^{n} -\epsilon(\alpha_i^+ + \alpha_i^-)\right)}$$

= ⟵ same as above ⟶

$$+\left(\sum_{i=1}^{n} y_i(\alpha_i^+ - \alpha_i^-)\right) + b\left[\sum_{i=1}^{n}\alpha_i^- - \sum_{i=1}^{n}\alpha_i^+\right] + \left(\sum_{i=1}^{n} -\epsilon(\alpha_i^+ + \alpha_i^-)\right)$$

= ⟵ same as above ⟶

$$+ (\text{same}) + b\left[\cancel{\sum_{i=1}^{n}\alpha_i^+} - \cancel{\sum_{i=1}^{n}\alpha_i^+}\right] + (\text{same})$$

$$L_D(\alpha^+,\alpha^-) = -\frac{1}{2}\sum_{i=1}^{n}(\alpha_i^+)^2\langle x_i,x_i\rangle - \sum_{i=1}^{n}(\alpha_i^+)(\alpha_i^-)\langle x_i,x_i\rangle + \frac{3}{2}\sum_{i=1}^{n}(\alpha_i^-)^2\langle x_i,x_i\rangle$$

$$+\left(\sum_{i=1}^{n} y_i(\alpha_i^+-\alpha_i^-)\right) + \cancel{b\left(\sum_{i=1}^{n}(\alpha_i^- + \alpha_i^+)\right)} - \epsilon\left(\sum_{i=1}^{n}(\alpha_i^+ + \alpha_i^-)\right)$$

- **Dual Optimization Problem:**

$$\max_{\alpha^+,\alpha^-,\beta^+,\beta^-}\left(L_D(\alpha^+,\alpha^-)\right) \text{ such that }$$

KKT conditions

$$\overbrace{\alpha_i^+ \geq 0,\ \alpha_i^- \geq 0,\ \beta_i^+ \geq 0,\ \beta_i^- \geq 0 \quad \forall i,}$$

$$\left.\begin{array}{l} \sum_{i=1}^{n}\alpha_i^+ - \sum_{i=1}^{n}\alpha_i^- = 0 \\ \frac{C}{n} - \alpha_i^+ - \beta_i^+ = 0 \\ \frac{C}{n} - \alpha_i^- - \beta_i^- = 0 \end{array}\right\} \nabla \text{ conditions}$$

Combine constraints

- $0 \leq \alpha_i^+ \leq \frac{C}{n} \quad \forall i$
- $0 \leq \alpha_i^- \leq \frac{C}{n} \quad \forall i$
- $\sum_{i=1}^{n}\alpha_i^+ - \sum_{i=1}^{n}\alpha_i^- = 0$

$-\frac{1}{2}x^2 - xy + \frac{3}{2}y^2$

- In summary, the dual optimization can be written as:

$$\max_{\alpha^+,\alpha^-} \Big[-\frac{1}{2}\sum_{i=1}^{n}(\alpha_i^+)^2\langle x_i, x_i\rangle - \sum_{i=1}^{n}(\alpha_i^+)(\alpha_i^-)\langle x_i, x_i\rangle + \frac{3}{2}\sum_{i=1}^{n}(\alpha_i^-)^2\langle x_i, x_i\rangle + \Big(\sum_{i=1}^{n} y_i(\alpha_i^+ - \alpha_i^-)\Big) \cancel{+ b\Big(\sum_{i=1}^{n}(\alpha_i^+ + \alpha_i^-)\Big)} - \epsilon\Big(\sum_{i=1}^{n}(\alpha_i^+ + \alpha_i^-)\Big) \Big]$$

such that

$$0 \le \alpha_i^+ \le C/n$$

$$0 \le \alpha_i^- \le C/n$$

$$\sum_{i=1}^{n}\alpha_i^+ - \sum_{i=1}^{n}\alpha_i^- = 0$$

Leave this unexpanded as in online solution.

SLDM hw 5, 2C : Explain how to kernelize SVR.
Be sure to explain how to recover $\underset{\sim}{w}^*$ and b^*

- Let $(\alpha^+)^*$ and $(\alpha^-)^*$ be dual optimal.
- From the KKT conditions we saw that $\underset{\sim}{w} = \sum_{i=1}^{n} x_i (\alpha_i^+ - \alpha_i^-)$.

There fore, $\underset{\sim}{w}^* = \sum_{i=1}^{n} x_i \left((\alpha_i^+)^* - (\alpha_i^-)^* \right)$.

- -- Consider and i such that $0 < (\alpha_i^+)^* < \frac{C}{n}$ and
$0 < (\alpha_i^-)^* < \frac{C}{n}$.

-- Since $(\alpha_i)^* > 0$ we know by the KKT conditions that

$$y_i - \langle \underset{\sim}{w}^*, x_i \rangle - b^* = \epsilon + \xi_i^+ \quad \text{and} \quad -y_i + \langle \underset{\sim}{w}^*, x_i \rangle + b^* = \epsilon + \xi_i^-.$$

-- Since $(\alpha_i^{+*}) < \frac{C}{n}$ by another KKT condition we get

$$\frac{C}{n} - (\alpha_i^*) = \beta_i^* > 0 \implies \xi_i^* = 0 \implies (\xi_i^+)^* = 0, \; (\xi_i^-)^* = 0.$$

- So now we can solve b^*:

$$b^* = y_i - \langle \underset{\sim}{w}^*, x_i \rangle - \epsilon - \xi_i^+$$
$$= y_i - \langle \underset{\sim}{w}^*, x_i \rangle - \epsilon - 0$$
$$= y_i - \sum_{j=1}^{n} (\alpha_j^{+*} - \alpha_j^{-*}) \langle x_j, x_i \rangle - \epsilon$$

2D on the back

- Similarly, when using α^{-*} we will solve for b^* and get

$b^* = y_i - \langle \underset{\sim}{w}^*, x_i \rangle - \epsilon + \xi_i^-$ where $\xi_i^- = 0$ giving

$$b^* = y_i - \sum_{j=1}^{n} (\alpha_j^{+*} - \alpha_j^{-*}) \langle x_j, x_i \rangle - \epsilon$$