

Exercise 4.10.1: The dual problem of the ν -SVR

$$(a) \min_{\underline{w}, b, \epsilon, \varphi_a, \varphi_a^*} \frac{1}{2} \|\underline{w}\|^2 + C \left[\nu \epsilon + \frac{1}{p} \sum_{a=1}^p (\varphi_a + \varphi_a^*) \right]$$

$$\begin{aligned} \text{s.t.} \quad & \underline{w}^T \underline{x}^{(a)} + b - y_T^{(a)} \leq \epsilon + \varphi_a \\ & y_T^{(a)} - \underline{w}^T \underline{x}^{(a)} - b \leq \epsilon + \varphi_a^* \\ & \varphi_a \geq 0 \\ & \varphi_a^* \geq 0 \\ & \epsilon \geq 0 \end{aligned} \quad \forall a \quad \begin{aligned} & \rightarrow \lambda_a \geq 0 \\ & \rightarrow \lambda_a^* \geq 0 \\ & \rightarrow \mu_a \geq 0 \\ & \rightarrow \mu_a^* \geq 0 \\ & \rightarrow \delta \geq 0 \end{aligned}$$

$$\begin{aligned} \mathcal{L}(\underline{w}, b, \epsilon, \varphi_a, \varphi_a^*, \lambda_a, \lambda_a^*, \mu_a, \mu_a^*, \delta) \\ = \frac{1}{2} \|\underline{w}\|^2 + C \left[\nu \epsilon + \frac{1}{p} \sum_{a=1}^p (\varphi_a + \varphi_a^*) \right] \\ + \sum_{a=1}^p \lambda_a (\underline{w}^T \underline{x}^{(a)} + b - y_T^{(a)} - \epsilon - \varphi_a) \\ + \sum_{a=1}^p \lambda_a^* (y_T^{(a)} - \underline{w}^T \underline{x}^{(a)} - b - \epsilon - \varphi_a^*) \\ - \sum_{a=1}^p \mu_a \varphi_a \\ - \sum_{a=1}^p \mu_a^* \varphi_a^* \\ - \delta \epsilon \end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial \underline{w}} = \underline{w} + \sum_{a=1}^p (\lambda_a - \lambda_a^*) \underline{x}^{(a)} \stackrel{!}{=} 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{a=1}^p (\lambda_a - \lambda_a^*) \stackrel{!}{=} 0$$

$$\frac{\partial \mathcal{L}}{\partial \epsilon} = C \nu - \sum_{a=1}^p (\lambda_a + \lambda_a^*) - \delta \stackrel{!}{=} 0$$

$$\frac{\partial \mathcal{L}}{\partial \varphi_a} = \frac{C}{p} - \lambda_a - \mu_a \stackrel{!}{=} 0$$

$$\frac{\partial \mathcal{L}}{\partial \varphi_a^*} = \frac{C}{p} - \lambda_a^* - \mu_a^* \stackrel{!}{=} 0$$

$$\underline{w} = \sum_{a=1}^p (\lambda_a^* - \lambda_a) \underline{x}^{(a)}$$

$$\sum_{a=1}^p (\lambda_a - \lambda_a^*) = 0$$

$$C \nu = \delta + \sum_{a=1}^p (\lambda_a + \lambda_a^*)$$

$$C = p(\lambda_a + \mu_a)$$

$$C = p(\lambda_a^* + \mu_a^*)$$

$$\begin{aligned}
 (b) \quad \mathcal{L} &= \frac{1}{2} \sum_{\alpha, \beta=1}^p (\lambda_\alpha^* - \lambda_\alpha) (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + C \nu \epsilon + \frac{C}{p} \sum_{\alpha=1}^p (\varphi_\alpha + \varphi_\alpha^*) \\
 &+ \sum_{\alpha, \beta=1}^p \lambda_\alpha (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + b \sum_{\alpha=1}^p \lambda_\alpha - \sum_{\alpha=1}^p \lambda_\alpha y_\tau^{(\alpha)} \\
 &- \epsilon \sum_{\alpha=1}^p \lambda_\alpha - \sum_{\alpha=1}^p \lambda_\alpha \varphi_\alpha + \sum_{\alpha=1}^p \lambda_\alpha^* y_\tau^{(\alpha)} - \sum_{\alpha, \beta=1}^p \lambda_\alpha^* (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} \\
 &- b \sum_{\alpha=1}^p \lambda_\alpha^* - \epsilon \sum_{\alpha=1}^p \lambda_\alpha^* - \sum_{\alpha=1}^p \lambda_\alpha^* \varphi_\alpha^* - \frac{C}{p} \sum_{\alpha=1}^p \varphi_\alpha + \sum_{\alpha=1}^p \lambda_\alpha \varphi_\alpha \\
 &- \frac{C}{p} \sum_{\alpha=1}^p \varphi_\alpha^* + \sum_{\alpha=1}^p \lambda_\alpha^* \varphi_\alpha^* - C \nu \epsilon + \epsilon \sum_{\alpha=1}^p (\lambda_\alpha + \lambda_\alpha^*)
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{L} &= \frac{1}{2} \sum_{\alpha, \beta=1}^p (\lambda_\alpha^* - \lambda_\alpha) (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} \\
 &+ \sum_{\alpha, \beta=1}^p (\lambda_\alpha - \lambda_\alpha^*) (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} \\
 &- \sum_{\alpha=1}^p \lambda_\alpha y_\tau^{(\alpha)} + \sum_{\alpha=1}^p \lambda_\alpha^* y_\tau^{(\alpha)}
 \end{aligned}$$

$$\max_{\lambda_\alpha, \lambda_\alpha^*} - \frac{1}{2} \sum_{\alpha, \beta=1}^p (\lambda_\alpha^* - \lambda_\alpha) (\lambda_\beta^* - \lambda_\beta) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \sum_{\alpha=1}^p (\lambda_\alpha^* - \lambda_\alpha) y_\tau^{(\alpha)}$$

and we have $\forall \alpha, \begin{matrix} \lambda_\alpha \geq 0 \\ \lambda_\alpha^* \geq 0 \\ \mu_\alpha \geq 0 \\ \mu_\alpha^* \geq 0 \\ \delta \geq 0 \end{matrix}$

$$\mu_\alpha = \frac{C}{p} - \lambda_\alpha \geq 0 \quad \Rightarrow \quad \frac{C}{p} \geq \lambda_\alpha$$

$$\mu_\alpha^* = \frac{C}{p} - \lambda_\alpha^* \geq 0 \quad \Rightarrow \quad \frac{C}{p} \geq \lambda_\alpha^*$$

$$\delta = C \nu - \sum_{\alpha=1}^p (\lambda_\alpha + \lambda_\alpha^*) \geq 0 \quad \Rightarrow \quad C \nu \geq \sum_{\alpha=1}^p (\lambda_\alpha + \lambda_\alpha^*)$$

so the constraints are:

$$\begin{aligned}
 \forall \alpha \in \{1, \dots, p\}: \quad & 0 \leq \lambda_\alpha \leq \frac{C}{p} & 0 \leq \lambda_\alpha^* \leq \frac{C}{p} \\
 & \sum_{\alpha=1}^p (\lambda_\alpha^* - \lambda_\alpha) = 0 & \sum_{\alpha=1}^p (\lambda_\alpha + \lambda_\alpha^*) \leq \nu C
 \end{aligned}$$