

### Exercise 4.10.1

The optimal  $\varepsilon$ -parameter depends on the noise level in the data, which is unknown. Derive the primal problem for the  $\nu$ -SVR, which adjusts  $\varepsilon$  as the primal parameter.

PRIMAL optimisation problem of  $\nu$ -SVR

$$\min_{\underline{w}, b, \varphi_L, \varphi_L^*, \varepsilon} \frac{1}{2} \|\underline{w}\|^2 + C \left[ \nu \varepsilon + \frac{1}{P} \sum_{L=1}^P (\varphi_L + \varphi_L^*) \right]$$

new variablenew parameter

subject to:

$$\begin{aligned} \underline{w}^T \underline{x}^{(L)} + b - \varphi_L^{(L)} &\leq \varepsilon + \varphi_L \\ \varphi_L^{(L)} - \underline{w}^T \underline{x}^{(L)} - b &\leq \varepsilon + \varphi_L^* \\ \varphi_L, \varphi_L^*, \varepsilon &\geq 0 \quad \forall L \in \overline{1, P} \end{aligned}$$

The Lagrangian of the  $\nu$ -SVR:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \|\underline{w}\|^2 + C \left[ \nu \varepsilon + \frac{1}{P} \sum_{L=1}^P (\varphi_L + \varphi_L^*) \right] - \sum_{L=1}^P \lambda_L \left[ \varepsilon + \varphi_L - \underline{w}^T \underline{x}^{(L)} - b + \varphi_L^{(L)} \right] \\ & - \sum_{L=1}^P \lambda_L^* \left[ \varepsilon + \varphi_L^* - \varphi_L^{(L)} + \underline{w}^T \underline{x}^{(L)} + b \right] - \sum_{L=1}^P \eta_L \varphi_L - \sum_{L=1}^P \eta_L^* \varphi_L^* - \delta \varepsilon \end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial \underline{w}} = \underline{w} + \sum_{L=1}^P \lambda_L \underline{x}^{(L)} - \sum_{L=1}^P \lambda_L^* \underline{x}^{(L)} = 0$$

$$\underline{w} = \sum_{L=1}^P (\lambda_L^* - \lambda_L) \underline{x}^{(L)} \quad (1)$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{L=1}^P \lambda_L - \sum_{L=1}^P \lambda_L^* = 0 \Rightarrow \sum_{L=1}^P (\lambda_L - \lambda_L^*) = 0 \quad (2)$$

$$\frac{\partial \mathcal{L}}{\partial \varphi_L} = \frac{C}{P} - \lambda_L - \eta_L = 0 \Rightarrow \eta_L = \frac{C}{P} - \lambda_L \quad (3)$$

$$\frac{\partial \mathcal{L}}{\partial \varphi_L^*} = \frac{C}{P} - \lambda_L^* - \eta_L^* = 0 \Rightarrow \eta_L^* = \frac{C}{P} - \lambda_L^* \quad (4)$$



$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = C\mathbf{v} - \sum_{L=1}^P \lambda_L - \sum_{L=1}^P \lambda_L^* - \delta = 0$$

$$\Rightarrow C\mathbf{v} - \sum_{L=1}^P (\lambda_L + \lambda_L^*) = \delta \quad (5)$$

$$\Rightarrow C\mathbf{v} = \sum_{L=1}^P (\lambda_L + \lambda_L^*) + \delta$$

If we substitute (1), (2), ..., (5) into the expression for the Lagrange primal we get the Lagrange dual:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{L=1}^P (\lambda_L^* - \lambda_L) \underline{x}^{(L)} \cdot \sum_{B=1}^P (\lambda_B^* - \lambda_B) \underline{x}^{(B)} + C\mathbf{v}^T \mathbf{E} + \frac{C}{P} \sum_{L=1}^P \psi_L + \frac{C}{P} \sum_{L=1}^P \psi_L^* \\ & - \sum_{L=1}^P \lambda_L \mathbf{E} - \sum_{L=1}^P \lambda_L \psi_L + \sum_{L=1}^P \lambda_L \sum_{B=1}^P (\lambda_B^* - \lambda_B) \underline{x}^{(B)} \underline{x}^{(L)} + \sum_{L=1}^P \lambda_L b \\ & - \sum_{L=1}^P \lambda_L \psi_L^{(L)} - \sum_{L=1}^P \lambda_L^* \mathbf{E} - \sum_{L=1}^P \lambda_L^* \psi_L^* + \sum_{L=1}^P \lambda_L^* \psi_L^{(L)} - \sum_{L=1}^P \lambda_L^* \sum_{B=1}^P (\lambda_B^* - \lambda_B) \underline{x}^{(B)} \underline{x}^{(L)} \\ & - \sum_{L=1}^P \lambda_L^* b - \sum_{L=1}^P \left( \frac{C}{P} - \lambda_L \right) \psi_L - \sum_{L=1}^P \left( \frac{C}{P} - \lambda_L^* \right) \psi_L^* - \left[ C\mathbf{v} - \sum_{L=1}^P (\lambda_L + \lambda_L^*) \right] \mathbf{E} \end{aligned}$$

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{L=1}^P \sum_{B=1}^P (\lambda_L^* - \lambda_L) (\lambda_B^* - \lambda_B) (\underline{x}^{(L)})^T \underline{x}^{(B)} + C\mathbf{v}^T \mathbf{E} + \frac{C}{P} \sum_{L=1}^P \psi_L + \frac{C}{P} \sum_{L=1}^P \psi_L^* \\ & - \sum_{L=1}^P \lambda_L \mathbf{E} - \sum_{L=1}^P \lambda_L \psi_L - \sum_{L=1}^P \sum_{B=1}^P (\lambda_L^* - \lambda_L) (\lambda_B^* - \lambda_B) (\underline{x}^{(L)})^T \underline{x}^{(B)} \\ & + \sum_{L=1}^P \lambda_L b - \sum_{L=1}^P \lambda_L \psi_L^{(L)} - \sum_{L=1}^P \lambda_L^* \mathbf{E} - \sum_{L=1}^P \lambda_L^* \psi_L^* + \sum_{L=1}^P \lambda_L^* \psi_L^{(L)} - \sum_{L=1}^P \lambda_L^* b \\ & - \frac{C}{P} \sum_{L=1}^P \psi_L + \sum_{L=1}^P \lambda_L \psi_L - \frac{C}{P} \sum_{L=1}^P \psi_L^* + \sum_{L=1}^P \lambda_L^* \psi_L^* - C\mathbf{v}^T \mathbf{E} + \mathbf{E} \sum_{L=1}^P (\lambda_L + \lambda_L^*) \end{aligned}$$

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2} \sum_{L,B=1}^P (\lambda_L^* - \lambda_L) (\lambda_B^* - \lambda_B) (\underline{x}^{(L)})^T \underline{x}^{(B)} - \sum_{L=1}^P (\lambda_L + \lambda_L^*) \mathbf{E} \\ & + \sum_{L=1}^P (\lambda_L - \lambda_L^*) b + \sum_{L=1}^P (\lambda_L^* - \lambda_L) \psi_L^{(L)} + \mathbf{E} \sum_{L=1}^P (\lambda_L + \lambda_L^*) \end{aligned}$$



$$\mathcal{L} = -\frac{1}{2} \sum_{\alpha, \beta=1}^P (\lambda_{\alpha}^* - \lambda_{\alpha}) (\lambda_{\beta}^* - \lambda_{\beta}) (\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \sum_{\alpha=1}^P (\lambda_{\alpha}^* - \lambda_{\alpha}) y_{\alpha}^{(\alpha)}$$

After eliminating all primal variable from the expression we get the Lagrange dual in a form:

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

$$\text{subject to: } \sum_{\alpha=1}^P (\lambda_{\alpha} - \lambda_{\alpha}^*) = 0$$

$$0 \leq \lambda_{\alpha}, \lambda_{\alpha}^* \leq \frac{C}{P}$$

$$\sum_{\alpha=1}^P (\lambda_{\alpha} + \lambda_{\alpha}^*) \leq \nu C$$