Exercise # 10.1

The optimal ε-parameter depends on the noise level in the data, which is unknown. Derive the primal problem for the ν-SVR, which adjusts ε as the primal parameter

PRIMAL optimization problem of ν-SVR

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

new variable (ε)

new parameter (ν)

subject to:

$$\underline{w}^T \underline{x}^{(\alpha)} + b - y_T^{(\alpha)} \le \varepsilon + \varphi_\alpha$$

$$y_T^{(\alpha)} - \underline{w}^T \underline{x}^{(\alpha)} - b \le \varepsilon + \varphi_\alpha^*$$

$$\varphi_\alpha, \varphi_\alpha^*, \varepsilon \ge 0 \qquad \forall \alpha \in \overline{1,p}$$

The Lagrangian of the ν-SVR:

$$\mathcal{L} = \frac{1}{2}\|w\|^2 + C\left[\nu\varepsilon + \frac{1}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi_\alpha^*)\right] - \sum_{\alpha=1}^{p}\lambda_\alpha\left[\varepsilon + \varphi_\alpha - \underline{w}^T\underline{x}^{(\alpha)} - b + y_T^{(\alpha)}\right]$$

$$- \sum_{\alpha=1}^{p}\lambda_\alpha^*\left[\varepsilon + \varphi_\alpha^* - y_T^{(\alpha)} + \underline{w}^T\underline{x}^{(\alpha)} + b\right] - \sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta_\alpha^*\varphi_\alpha^* - \delta\varepsilon$$

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

$$w = \sum_{\alpha=1}^{p}(\lambda_\alpha^* - \lambda_\alpha)\underline{x}^{(\alpha)} \qquad (1)$$

$$\frac{\partial\mathcal{L}}{\partial b} = \sum_{\alpha=1}^{p}\lambda_\alpha - \sum_{\alpha=1}^{p}\lambda_\alpha^* = 0 \Rightarrow \sum_{\alpha=1}^{p}(\lambda_\alpha - \lambda_\alpha^*) = 0 \qquad (2)$$

$$\frac{\partial\mathcal{L}}{\partial\varphi_\alpha} = \frac{C}{p} - \lambda_\alpha - \eta_\alpha = 0 \Rightarrow \eta_\alpha = \frac{C}{p} - \lambda_\alpha \qquad (3)$$

$$\frac{\partial\mathcal{L}}{\partial\varphi_\alpha^*} = \frac{C}{p} - \lambda_\alpha^* - \eta_\alpha^* = 0 \Rightarrow \eta_\alpha^* = \frac{C}{p} - \lambda_\alpha^* \qquad (4)$$

$$\frac{\partial \mathcal{L}}{\partial \varepsilon} = C\nu - \sum_{\alpha=1}^{p} \lambda_\alpha - \sum_{\alpha=1}^{p} \lambda_\alpha^* - \delta = 0 \Rightarrow C\nu = \sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*) + \delta$$

$$\Rightarrow \underline{C\nu - \sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*) = \delta} \quad (5)$$

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

$$\mathcal{L} = \frac{1}{2}\sum_{\alpha=1}^{p} (\lambda_\alpha^* - \lambda_\alpha)\, \underline{x}^{(\alpha)} \cdot \sum_{\beta=1}^{p} (\lambda_\beta^* - \lambda_\beta)\, \underline{x}^{(\beta)} + C\nu\varepsilon + \frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha + \frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha^*$$

$$- \sum_{\alpha=1}^{p} \lambda_\alpha \varepsilon - \sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha + \sum_{\alpha=1}^{p} \lambda_\alpha \sum_{\beta=1}^{p} (\lambda_\beta^* - \lambda_\beta)\, \underline{x}^{(\beta)}\, \underline{x}^{(\alpha)} + \sum_{\alpha=1}^{p} \lambda_\alpha b$$

$$- \sum_{\alpha=1}^{p} \lambda_\alpha y_T^{(\alpha)} - \sum_{\alpha=1}^{p} \lambda_\alpha^* \varepsilon - \sum_{\alpha=1}^{p} \lambda_\alpha^* \varphi_\alpha^* + \sum_{\alpha=1}^{p} \lambda_\alpha^* y_T^{(\alpha)} - \sum_{\alpha=1}^{p} \lambda_\alpha^* \sum_{\beta=1}^{p} (\lambda_\beta^* - \lambda_\beta)\, \underline{x}^{(\beta)}\, \underline{x}^{(\alpha)}$$

$$- \sum_{\alpha=1}^{p} \lambda_\alpha^* b - \sum_{\alpha=1}^{p} \left(\frac{C}{p} - \lambda_\alpha\right) \varphi_\alpha - \sum_{\alpha=1}^{p} \left(\frac{C}{p} - \lambda_\alpha^*\right) \varphi_\alpha^* - \left[C\nu - \sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*)\right]\varepsilon$$

$$\mathcal{L} = \frac{1}{2}\sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} (\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)\left(\underline{x}^{(\alpha)}\right)^T \underline{x}^{(\beta)} + \cancel{C\nu\varepsilon} + \cancel{\frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha} + \cancel{\frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha^*}$$

$$- \sum_{\alpha=1}^{p} \lambda_\alpha \varepsilon - \cancel{\sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha} - \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} (\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)\left(\underline{x}^{(\alpha)}\right)^T \underline{x}^{(\beta)}$$

$$+ \sum_{\alpha=1}^{p} \lambda_\alpha b - \sum_{\alpha=1}^{p} \lambda_\alpha y_T^{(\alpha)} - \sum_{\alpha=1}^{p} \lambda_\alpha^* \varepsilon - \cancel{\sum_{\alpha=1}^{p} \lambda_\alpha^* \varphi_\alpha^*} + \sum_{\alpha=1}^{p} \lambda_\alpha^* y_T^{(\alpha)} - \sum_{\alpha=1}^{p} \lambda_\alpha^* b$$

$$- \cancel{\frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha} + \cancel{\sum_{\alpha=1}^{p} \lambda_\alpha \varphi_\alpha} - \cancel{\frac{C}{p}\sum_{\alpha=1}^{p} \varphi_\alpha^*} + \cancel{\sum_{\alpha=1}^{p} \lambda_\alpha^* \varphi_\alpha^*} - \cancel{C\nu\varepsilon} + \varepsilon\sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*)$$

$$\mathcal{L} = -\frac{1}{2}\sum_{\alpha,\beta=1}^{p} (\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)\left(\underline{x}^{(\alpha)}\right)^T \cdot \underline{x}^{(\beta)} - \cancel{\sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*)\varepsilon}$$

$$+ \overbrace{\sum_{\alpha=1}^{p} (\lambda_\alpha - \lambda_\alpha^*) b}^{0\ (2)} + \sum_{\alpha=1}^{p} (\lambda_\alpha^* - \lambda_\alpha) y_T^{(\alpha)} + \cancel{\varepsilon\sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*)}$$

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

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

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

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

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

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