Machine Intelligence 1 - Exercise 10: Support Vector Regression

Liu, Zhiwei 387571

Moon, Chulhyun 392865

Ozmen, Cengizhan 388011

Pipo, Aiko 390011

January 24, 2018

H10.1: The dual problem of the ν-SVR (5 points)

(a)

We have the Lagrangian of the ν-SVR given as

$$\begin{aligned}
&\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi^*_\alpha\}, \epsilon, \{\lambda_\alpha\}, \{\lambda^*_\alpha\}, \{\eta_\alpha\}, \{\eta^*_\alpha\}, \delta) = \\
&\frac{1}{2}||\underline{w}||^2 + C\left[\nu\epsilon + \frac{1}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha)\right] \\
&-\sum_{\alpha=1}^{p}\lambda_\alpha(\varphi_\alpha + \epsilon + y_T^{(\alpha)} - \underline{w}^T\underline{x}^{(\alpha)} - b) \\
&-\sum_{\alpha=1}^{p}\lambda^*_\alpha(\varphi^*_\alpha + \epsilon + \underline{w}^T\underline{x}^{(\alpha)} + b - y_T^{(\alpha)}) \\
&-\sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta^*_\alpha\varphi^*_\alpha - \delta\epsilon
\end{aligned}$$

with the primal parameters

$$\underline{w}, b, \{\varphi_\alpha\}, \{\varphi^*_\alpha\}, \epsilon.$$

The partial derivatives w.r.t. the primal parameters are actually quite easy to compute:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\underline{w}} &= \frac{\partial}{\partial\underline{w}}\left(\frac{1}{2}||\underline{w}||^2 - \sum_{\alpha=1}^{p} -\lambda_\alpha\underline{w}^T\underline{x}^{(\alpha)} - \sum_{\alpha=1}^{p}\lambda^*_\alpha\underline{w}^T\underline{x}^{(\alpha)}\right) \\
&= \underline{w} - \sum_{\alpha=1}^{p}\frac{\partial}{\partial\underline{w}}(\lambda^*_\alpha\underline{w}^T\underline{x}^{(\alpha)} - \lambda_\alpha\underline{w}^T\underline{x}^{(\alpha)}) \\
&= \underline{w} - \sum_{\alpha=1}^{p}\lambda^*_\alpha\underline{x}^{(\alpha)} - \lambda_\alpha\underline{x}^{(\alpha)} \\
&= \underline{w} - \sum_{\alpha=1}^{p}(\lambda^*_\alpha - \lambda_\alpha)\underline{x}^{(\alpha)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial b} &= \frac{\partial}{\partial b}\left(-\sum_{\alpha=1}^{p} -\lambda_\alpha b - \sum_{\alpha=1}^{p}\lambda^*_\alpha b\right) \\
&= \sum_{\alpha=1}^{p}\lambda_\alpha - \lambda^*_\alpha
\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\varphi_\alpha} &= \frac{\partial}{\partial\varphi_\alpha}\left(\frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha+\varphi_\alpha^*) - \sum_{\alpha=1}^{p}\lambda_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha\right)\\ &= \frac{C}{p} - \lambda_\alpha - \eta_\alpha\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\varphi_\alpha^*} &= \frac{\partial}{\partial\varphi_\alpha^*}\left(\frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha+\varphi_\alpha^*) - \sum_{\alpha=1}^{p}\lambda_\alpha^*\varphi_\alpha^* - \sum_{\alpha=1}^{p}\eta_\alpha^*\varphi_\alpha^*\right)\\ &= \frac{C}{p} - \lambda_\alpha^* - \eta_\alpha^*\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\epsilon} &= \frac{\partial}{\partial\epsilon}\left(C\nu\epsilon - \sum_{\alpha=1}^{p}\lambda_\alpha\epsilon - \sum_{\alpha=1}^{p}\lambda_\alpha^*\epsilon - \delta\epsilon\right)\\ &= C\nu - \sum_{\alpha=1}^{p}(\lambda_\alpha+\lambda_\alpha^*) - \delta\end{aligned}$$

(b)

By setting all derivatives above to zero, we got:

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

$$\sum_{\alpha=1}^{p}(\lambda_\alpha^* - \lambda_\alpha) = 0$$

$$\lambda_\alpha + \eta_\alpha = \frac{C}{p}$$

$$\lambda_\alpha^* + \eta_\alpha^* = \frac{C}{p}$$

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

Now reorder the Lagrangian:

$$\begin{aligned}
&\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi^*_\alpha\}, \epsilon, \{\lambda_\alpha\}, \{\lambda^*_\alpha\}, \{\eta_\alpha\}, \{\eta^*_\alpha\}, \delta) = \\
&\frac{1}{2}||\underline{w}||^2 + C\nu\epsilon + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha) \\
&- \sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)\underline{w}^T\underline{x}^{(\alpha)} \\
&+ \sum_{\alpha=1}^{P}(\lambda_\alpha - \lambda^*_\alpha)b \\
&+ \sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)} \\
&- \sum_{\alpha=1}^{P}(\lambda_\alpha + \lambda^*_\alpha)\epsilon \\
&- \sum_{\alpha=1}^{P}\lambda_\alpha\varphi_\alpha - \sum_{\alpha=1}^{P}\lambda^*_\alpha\varphi^*_\alpha \\
&- \sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta^*_\alpha\varphi^*_\alpha - \delta\epsilon
\end{aligned}$$

put the results we got from above into the Lagrangian:

$$\begin{aligned}
&\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi^*_\alpha\}, \epsilon, \{\lambda_\alpha\}, \{\lambda^*_\alpha\}, \{\eta_\alpha\}, \{\eta^*_\alpha\}, \delta) = \\
&\frac{1}{2}||\underline{w}||^2 + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha) \\
&- \sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)\underline{w}^T\underline{x}^{(\alpha)} \\
&+ \sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)} \\
&- \sum_{\alpha=1}^{P}\lambda_\alpha\varphi_\alpha - \sum_{\alpha=1}^{P}\lambda^*_\alpha\varphi^*_\alpha \\
&- \sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta^*_\alpha\varphi^*_\alpha
\end{aligned}$$

then put $\underline{w}$ into it:

$$
\begin{aligned}
&\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi_\alpha^*\}, \epsilon, \{\lambda_\alpha\}, \{\lambda_\alpha^*\}, \{\eta_\alpha\}, \{\eta_\alpha^*\}, \delta) = \\
&\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)((\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi_\alpha^*) \\
&-\sum_{\alpha,\beta=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)\underline{x}^{(\alpha)}(\lambda_\beta^* - \lambda_\beta)((\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} \\
&+\sum_{\alpha=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)y_T^{(\alpha)} \\
&-\sum_{\alpha=1}^{P}\lambda_\alpha\varphi_\alpha - \sum_{\alpha=1}^{P}\lambda_\alpha^*\varphi_\alpha^* \\
&-\sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta_\alpha^*\varphi_\alpha^* \\
=\\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)((\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi_\alpha^*) \\
&+\sum_{\alpha=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)y_T^{(\alpha)} \\
&-\sum_{\alpha=1}^{P}\lambda_\alpha\varphi_\alpha - \sum_{\alpha=1}^{P}\lambda_\alpha^*\varphi_\alpha^* \\
&-\sum_{\alpha=1}^{p}\eta_\alpha\varphi_\alpha - \sum_{\alpha=1}^{p}\eta_\alpha^*\varphi_\alpha^* \\
=\\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)((\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi_\alpha^*) \\
&+\sum_{\alpha=1}^{P}(\lambda_\alpha^* - \lambda_\alpha)y_T^{(\alpha)} \\
&-\sum_{\alpha=1}^{P}(\lambda_\alpha + \eta_\alpha)\varphi_\alpha \\
&-\sum_{\alpha=1}^{P}(\lambda_\alpha^* + \eta_\alpha^*)\varphi_\alpha^*
\end{aligned}
$$

since we have known that

$$\lambda_\alpha + \eta_\alpha = \frac{C}{p}$$

$$\lambda_\alpha^* + \eta_\alpha^* = \frac{C}{p}$$

So the Lagrangian becomes:

$$
\begin{aligned}
&\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi^*_\alpha\}, \epsilon, \{\lambda_\alpha\}, \{\lambda^*_\alpha\}, \{\eta_\alpha\}, \{\eta^*_\alpha\}, \delta) = \\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)(\lambda^*_\beta - \lambda_\beta)((\underline{x}^{(\alpha)})^T\underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha) \\
&+\sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)} \\
&-\sum_{\alpha=1}^{P}\frac{C}{p}\varphi_\alpha \\
&-\sum_{\alpha=1}^{P}\frac{C}{p}\varphi^*_\alpha \\
=& \\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)(\lambda^*_\beta - \lambda_\beta)((\underline{x}^{(\alpha)})^T\underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha) \\
&+\sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)} \\
&-\frac{C}{p}(\sum_{\alpha=1}^{P}\varphi_\alpha + \sum_{\alpha=1}^{P}\varphi^*_\alpha) \\
=& \\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)(\lambda^*_\beta - \lambda_\beta)((\underline{x}^{(\alpha)})^T\underline{x}^{(\beta)} + \frac{C}{p}\sum_{\alpha=1}^{p}(\varphi_\alpha + \varphi^*_\alpha) \\
&+\sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)} \\
&-\frac{C}{p}\sum_{\alpha=1}^{P}(\varphi_\alpha + \varphi^*_\alpha) \\
=& \\
&-\frac{1}{2}\sum_{\alpha,\beta=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)(\lambda^*_\beta - \lambda_\beta)((\underline{x}^{(\alpha)})^T\underline{x}^{(\beta)} + \sum_{\alpha=1}^{P}(\lambda^*_\alpha - \lambda_\alpha)y_T^{(\alpha)}
\end{aligned}
$$

Now about the constraints of the λ_α and λ^*_α, we have already known that:

$$\lambda_\alpha + \eta_\alpha = \frac{C}{p}$$

$$\sum_{\alpha=1}^{p}(\lambda^*_\alpha - \lambda_\alpha) = 0$$

$$\lambda^*_\alpha + \eta^*_\alpha = \frac{C}{p}$$

$$\sum_{\alpha=1}^{p}(\lambda_\alpha + \lambda^*_\alpha) = C\nu - \delta$$

since:

$$\eta_\alpha \geq 0, \qquad \eta_\alpha^* \geq 0, \qquad \delta_\alpha \geq 0$$

so we have:

$$0 \leq \lambda_\alpha \leq \frac{C}{p}, \qquad 0 \leq \lambda_\alpha^* \leq \frac{C}{p}, \qquad \sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*) \leq C\nu$$

So to sum up, we could say:

$$\begin{gathered}\mathcal{L}(\underline{w}, b, \{\varphi_\alpha\}, \{\varphi_\alpha^*\}, \epsilon, \{\lambda_\alpha\}, \{\lambda_\alpha^*\}, \{\eta_\alpha\}, \{\eta_\alpha^*\}, \delta) = \\ -\frac{1}{2} \sum_{\alpha,\beta=1}^{P} (\lambda_\alpha^* - \lambda_\alpha)(\lambda_\beta^* - \lambda_\beta)((\underline{x}^{(\alpha)})^T \underline{x}^{(\beta)} + \sum_{\alpha=1}^{P} (\lambda_\alpha^* - \lambda_\alpha) y_T^{(\alpha)}\end{gathered}$$

w.r.t:

$$0 \leq \lambda_\alpha \leq \frac{C}{p}, \qquad 0 \leq \lambda_\alpha^* \leq \frac{C}{p}, \qquad \sum_{\alpha=1}^{p} (\lambda_\alpha^* - \lambda_\alpha) = 0, \qquad \sum_{\alpha=1}^{p} (\lambda_\alpha + \lambda_\alpha^*) \leq C\nu$$