7.11 The hinge loss (which is the loss used by the support vector machine) is given by

$$L(\alpha) = max\{0, 1-\alpha\},$$

If we are interested in applying gradient methods such as L-BFGS, and do not want to resort to subgradient methods, we need to smooth the kink in the hinge loss. Compute the convex conjugate of the hinge loss $L^\star(\beta)$ where β is the dual variable. Add a ℓ_2 proximal term, and compute the conjugate of the resulting function

$$L^\star(\beta) + \frac{\gamma}{2}\beta^2,$$

where γ is a given hyperparameter.

Solution.

$$L(\alpha) = \begin{cases} 1-\alpha, & \text{if } \alpha < 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

From Definition 7.4,

$$L^\star(\beta) = \sup_\alpha((\beta\alpha) - max\{0, 1-\alpha\})$$

$$L^\star(\beta) = \begin{cases} \sup\limits_\alpha((\beta\alpha) - (1-\alpha)), & \text{if } \alpha < 1 \\ \sup\limits_\alpha(\beta\alpha - 0), & \text{otherwise} \end{cases} \tag{2}$$

$$L^\star(\beta) = \begin{cases} \sup\limits_{\alpha}(\beta\alpha - 1 + \alpha), & \text{if } \alpha < 1 \\ \sup\limits_{\alpha}(\beta\alpha), & \text{otherwise} \end{cases} \tag{3}$$

$$L^\star(\beta) = \begin{cases} \sup\limits_{\alpha}((\beta + 1)\alpha - 1), & \text{if } \alpha < 1 \\ \sup\limits_{\alpha}(\beta\alpha), & \text{otherwise} \end{cases} \tag{4}$$

$$L^\star(\beta) = \begin{cases} \beta, & \text{if } \beta \geq -1, \alpha < 1 \\ \infty, & \text{if } \beta < -1, \alpha < 1 \\ \infty, & \text{if } \beta > 0, \alpha \geq 1 \\ \beta, & \text{if } \beta \leq 0, \alpha \geq 1 \end{cases} \tag{5}$$

$$\implies L^\star(\beta) = \begin{cases} \beta, & \text{if } 0 \geq \beta \geq -1 \\ \infty, & \text{otherwise} \end{cases} \tag{6}$$

Removing the infinity part, the conjugate with the proximal term gives us:

$$L_\gamma^\star(\beta) = \left(\beta + \frac{\gamma}{2}\beta^2\right) \qquad \text{where } 0 \geq \beta \geq -1$$

From Definition 7.4,

$$L_\gamma^{\star\star}(\alpha) = \sup_{\beta} \quad \left(\alpha\beta - (\beta + \frac{\gamma}{2}\beta^2)\right) \qquad \text{where } 0 \geq \beta \geq -1$$

$$L_\gamma^{\star\star}(\alpha) = \sup_{\beta} \quad \left(\beta(\alpha - 1 - \frac{\gamma}{2}\beta)\right) \qquad \text{where } 0 \geq \beta \geq -1$$

Taking derivative w.r.t β and setting it to 0, we get:

$$\alpha - 1 - \gamma\beta = 0$$

$$\implies \beta = \frac{(\alpha - 1)}{\gamma}$$

The second derivative $= -\gamma$.

If $\gamma > 0$, we get a local maxima at $\beta_{max} = \frac{(\alpha-1)}{\gamma}$.
We substitute this value for β back, to get the maxima:

$$\frac{(\alpha - 1)}{\gamma}\left(\alpha - 1 - \frac{\gamma}{2}\frac{(\alpha - 1)}{\gamma}\right)$$

$$= \frac{(\alpha - 1)}{\gamma}\left(\alpha - 1 - \frac{(\alpha - 1)}{2}\right)$$

$$= \frac{(\alpha - 1)^2}{2\gamma}$$

This local maxima can be located in 3 areas:

- $-1 \leq \beta_{max} \leq 0$

  The range of values of α can be found:

$$0 \geq \beta \geq -1$$

$$\implies 0 \geq \frac{(\alpha - 1)}{\gamma} \geq -1$$

$$\implies 0 \geq (\alpha - 1) \geq -\gamma$$

$$\implies 1 \geq \alpha \geq 1 - \gamma$$

$$\implies \text{If } \gamma > 0, \quad L_\gamma^{\star\star}(\alpha) = \frac{(\alpha-1)^2}{2\gamma} \quad \text{where } 1 \geq \alpha \geq 1-\gamma$$

- $\beta_{max} < -1$

  Function is decreasing from $\beta = -1$ to $\beta = 0$, so the supernum exists at $\beta = -1$.

  $\beta_{max} < -1 \quad \implies \quad L_\gamma^{\star\star}(\alpha) = \left(-1(\alpha - 1 - \frac{\gamma}{2}(-1))\right) = 1 - \alpha - \frac{\gamma}{2}$

  The range of values for α can be found:

$$\frac{(\alpha-1)}{\gamma} < -1$$

$$\implies \alpha < 1-\gamma$$

- $\beta_{max} > 0$

  Function is decreasing from $\beta = 0$ to $\beta = -1$, so the supernum exists at $\beta = 0$.

  $\beta_{max} < -1 \quad \implies \quad L_\gamma^{\star\star}(\alpha) = \left(0(\alpha - 1 - \frac{\gamma}{2}(0))\right) = 0$

  The range of values for α can be found:

$$\frac{(\alpha-1)}{\gamma} > 0$$

$$\implies \alpha > 1$$

$$\implies \text{If } \gamma > 0, \quad L_\gamma^{\star\star}(\alpha) = \begin{cases} 1 - \alpha - \frac{\gamma}{2}, & \text{if } \alpha < 1 - \gamma \\ \frac{(\alpha-1)^2}{2\gamma}, & \text{if } 1 \geq \alpha \geq 1 - \gamma \\ 0, & \text{if } \alpha > 1 \end{cases}$$

If $\gamma < 0$, we get a local minima at $\beta = \frac{(\alpha-1)}{\gamma}$. This means, the value of the function will increase as we move away from this point, and since there is only 1 minima, the maximum values will be encountered at the extremes of β.

$$\text{At } \beta = 0, \quad L_\gamma^\star(\beta) = 0$$

$$\text{and at } \beta = -1, \quad L_\gamma^\star(\beta) = \left(-1(\alpha - 1 - \frac{\gamma}{2}(-1))\right) = 1 - \alpha - \frac{\gamma}{2}$$

$$\implies \text{If } \gamma < 0, \quad L_\gamma^{\star\star}(\alpha) = \begin{cases} 0, & \text{if } \alpha \geq 1 - \frac{\gamma}{2} \\ 1 - \alpha - \frac{\gamma}{2}, & \text{otherwise} \end{cases}$$

$$\text{If } \gamma = 0,$$

$$L_\gamma^{\star\star}(\alpha) = \sup_\beta \quad (\beta(\alpha - 1)) \qquad \text{where } 0 \geq \beta \geq -1$$

$$\implies \text{If } \gamma = 0, \quad L_\gamma^{\star\star}(\alpha) = \begin{cases} 0, & \text{if } \alpha \geq 1 \\ 1 - \alpha, & \text{otherwise} \end{cases}$$