Problem 2: Soft Margin Hyperplanes [9 points]

The soft-margin primal SVM problem is:

$$\min \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{N} \xi_i$$

subject to feasibility constraints that for all $i = 1, ..., N$:

$$y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0$$

Suppose instead the optimization objective was changed to $\frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{N} \xi_i^p$ with $p > 1$ while the feasibility constraints are kept the same.

Hint: Before attempting this problem, we highly recommend first trying to derive the dual formulations for the hard-margin and soft-margin SVMs presented in class by working through the steps detailed below.

1. **[1 pt]** Incorporate the feasibility constraints into the objective function using the method of Lagrange multipliers by writing out the Lagrangian and the associated constraints.

$$\min \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N} \varepsilon_i^p \Rightarrow \begin{cases} \min \frac{1}{2}||w||_2^2 \\ \min C\sum_{i=1}^{N} \varepsilon_i^p \end{cases} \left(\begin{array}{c} \text{with constraints} \\ y^{(i)}(w^Tx^{(i)}+b) \geq 1-\varepsilon_i \\ \varepsilon_i \geq 0 \end{array} \right)$$

$$\mathcal{L}(w, b, \alpha, \varepsilon) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{N} \varepsilon_i^p - \sum_{i=1}^{N}\alpha_i\left(y^{(i)}(w^Tx^{(i)}+b)-1+\varepsilon_i\right)$$
$$= \frac{w^Tw}{2} + C\sum_{i=1}^{N} \varepsilon_i^p - \sum_{i=1}^{N}\alpha_i\left(y^{(i)}(w^Tx^{(i)}+b)-1+\varepsilon_i\right)$$

$\Rightarrow$ objective function

$$\min_{w,b,\varepsilon} \max_{\alpha} \frac{w^Tw}{2} + C\sum_{i=1}^{N} \varepsilon_i^p - \sum_{i=1}^{N}\alpha_i\left(y^{(i)}(w^Tx^{(i)}+b)-1+\varepsilon_i\right)$$

2. **[3 pts]** Next, take the partial derivative of the Lagrangian with respect to all of the optimization variables. Be sure to show all of you work.

$$\mathcal{L}(w, b, \alpha, \varepsilon) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} \varepsilon_i^P - \sum_{i=1}^{N}\alpha_i\left(y^{(i)}(w^T x^{(i)} + b) - 1 + \varepsilon_i\right)$$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2} \times 2w + 0 - \sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)} + 0 + 0 + 0$$

$$= w - \sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 + 0 - \sum_{i=1}^{N} 0 + \alpha_i y^{(i)} + 0 + 0$$

$$= -\sum_{i=1}^{N} \alpha_i y^{(i)}$$

$$\frac{\partial \mathcal{L}}{\partial \varepsilon_i} = 0 + cP\varepsilon_i^{P-1} - \alpha_i(0 + 0 + 0 + 1)$$

$$= cP\varepsilon_i^{P-1} - \alpha_i$$

3. **[3 pts]** Finally, by setting the partial derivatives from the previous part equal to 0, derive the dual formulation of SVMs in this general case. Again, be sure to show all of you work.

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)} = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{N} \alpha_i y^{(i)} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \varepsilon_i} = CP\varepsilon_i^{P-1} - \alpha_i = 0 \Rightarrow \alpha_i = CP\varepsilon_i^{P-1}$$

$$\Rightarrow \varepsilon_i = \left(\frac{\alpha_i}{CP}\right)^{\frac{1}{P-1}}$$

Substitutin w, ε_i to $\mathcal{L}$, the objective fuction

$$\Rightarrow \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} \varepsilon_i^P - \sum_{i=1}^{N} \alpha_i \left(y^{(i)}(w^T x^{(i)} + b) - 1 + \varepsilon_i\right)$$

$$= \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N} \left(\frac{\alpha_i}{CP}\right)^{\frac{P}{P-1}} - \sum_{i=1}^{N} \alpha_i \left(y^{(i)}\left(\sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)}\right) x^{(i)} + b\right) - 1 + \varepsilon_i\Big)$$

$$= \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{N}\left[C\left(\frac{\alpha_i}{CP}\right)^{\frac{P}{P-1}} - CP\varepsilon^P\right] + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i y^{(i)} b \underbrace{- \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}}_{\|w\|_2^2}$$

$$= \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N}\left[C\left(\frac{\alpha_i}{CP}\right)^{\frac{P}{P-1}} - CP\left(\frac{\alpha_i}{CP}\right)^{\frac{P}{P-1}}\right] - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$\Rightarrow \max_{\alpha} \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N}\left[C\left(\frac{\alpha_i}{CP}\right)^{\frac{P}{P-1}}(1-P)\right] - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

#

$(0 \le \alpha_i \le CP\varepsilon_i^{P-1})$