867002105_HW5

Charles Dotson

October 3, 2022

Contents

1 Let the Lagrangian function of the primal SVM be

$$L(w, \lambda) = f(w) - \sum_{i=1}^{m} \lambda_i g_i(w) = f(w) - \lambda^T g(w)$$

with the multipliers $\lambda \succeq 0$ and inequality constraints $g(w) \succeq 0$. Prove the weak duality property:

$$\max_{\lambda \succeq 0} \min_{w} L(w, \lambda) \leq \min_{w} \max_{\lambda \succeq 0} L(w, \lambda)$$

We are given the following primal problem:

$$\begin{aligned} \max \quad & L(w, \lambda) = f(w) - \lambda^T g(w) \\ \text{s.t.} \quad & \lambda \succeq 0 \end{aligned}$$

We set $\lambda^{'}$ to be the optimal solution to the dual. Plugging this into the same function where all else is equal. We can clearly see that the subtractive term could become positive at certain W. This would then yeild the result $L(w, \lambda^{'}) \leq \max_{\lambda \succeq 0} L(w, \lambda)$

We now add in the primal constraint. $\min_w L(w, \lambda^{'}) \leq \min_w \max_{\lambda \succeq 0} L(w, \lambda)$. To prove this one simply needs to see that we are maximizing the objective without the primal constraint and this will yield $\min_w L(w)$ as it is not a function of λ while $\lambda^{'}$ is the optimal soution already as well as minimizing a convex function. We can clearly see that the first relationship holds. These inequalities show that $\max_{\lambda \succeq 0} \min_w L(w, \lambda) \leq \min_w \max_{\lambda \succeq 0} L(w, \lambda)$

2 Question 2

We start by using the hyperplane $y = x$ in point slope form. Finding the $w^{\perp}$ to satisfy $w^T x = 0$ will give us \$ w = [-1, 1]^T\$. We now set up our maximum margin linear classifier $\widehat{h}(x; w)$. This classifier will also be used to have the new values used for the geometric margin calculations $(y^{(i)})$. Also, through linear algebra, one can see that the perpendicular distance aka the margin, is $\frac{\sqrt{2}}{2}$.

$$h(x; w) = \begin{cases} 1 & \text{if } -x_1^{(i)} + x_2^{(i)} > 0 \\ -1 & \text{if } -x_1^{(i)} + x_2^{(i)} \leq 0 \end{cases}$$

We can now see that, based on our new classifier, that $y^{(1)}, y^{(2)} = 1, -1$ respectively. The change creates a correct margin calulation as well as properly classiying the trainging examples. Shown below:

$$\frac{y^{(1)}(w_2)}{\|w\|} = \frac{\sqrt{2}}{2} = \frac{y^{(2)}(w_1)}{\|w\|}$$

$$\frac{y^{(1)}}{\sqrt{2}} = \frac{\sqrt{2}}{2} = \frac{-y^{(2)}}{\sqrt{2}}$$

$$\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$$

This shows we have correctly formulated the classifier for our training example.

3 Question 3

$$L(h) = (E)[\mathbb{1}[h(x) \neq (x)]]$$

$$= P\left(\frac{2\pi}{3} < \theta < \frac{5\pi}{4}\right) + P\left(\frac{5\pi}{4} < \theta < \frac{7\pi}{4}\right)$$

$$= \left(\frac{\pi}{2}\right)\frac{1}{2\pi} + \left(\frac{\pi}{4}\right)\frac{1}{2\pi}$$

$$= \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$$

4 Question 4

$$L(x : \{x^{(i)}, y^{(i)}\}_{i=1}^{m}) = \frac{1}{2}\sum_{i=1}^{m}(y^{(i)} - \langle w, x^{(i)}\rangle)^2$$

$$= \frac{1}{2}\left[\sum_{i=1}^{m}(y^{(i)})^2 - 2\sum_{i=1}^{m} y^{(i)}\langle w, x^{(i)}\rangle + \sum_{i=1}^{m}\langle w, x^{(i)}\rangle^2\right]$$

$$= \frac{1}{2}\left[\sum_{i=1}^{m}(y^{(i)})^2 - 2\sum_{i=1}^{m}\sum_{j=1}^{m} y^{(i)}\alpha_j K_{ij} + \left(\sum_{i=1}^{m}\alpha_j K_{ij}\right)^2\right]$$

5 Question 5