867002105_HW4

Charles Dotson

September 25, 2022

Contents

1 Given the two training examples $x^{(1)} = [2,1]^T$, $y^{(1)} = 1$ and $x(2) = [1,-1]^T$, $y^{(2)} = -1$, First write down the SVM primal problem and plug in the data (so that w and b are variables). Then construct the Lagrangian function $L(w,b,\lambda)$.

1.1 The Primal Formulation

$$\min_{w,b} \quad f(w) = \frac{1}{2}\|w\|^2$$

$$\text{s.t} \quad \left[w^T \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b\right] - 1 \geq 0$$

$$-\left[w^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} + b\right] - 1 \geq 0$$

1.2 Lagrangian Function $L(w,b,\lambda)$

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

$$= \frac{1}{2}\|w\|^2 - \sum_{i-1}^{m} \lambda_i (y^{(i)}(w^T x^{(i)} + b) - 1)$$

$$= \frac{1}{2}\|w\|^2 - \left[\lambda_1 \left[\left[w^T \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b\right] - 1\right] - \lambda_2 \left[\left[w^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} + b\right] - 1\right]\right]$$

2 Write down the KKT conditions of the above problem. Simplify the partial derivatives and write down the elements of the vectors (i.e., don't vectorize the conditions).

2.1 KKT Conditions

$$\frac{1}{2} - \lambda_1 x^{(1)} + \lambda_2 x^{(2)} = 0$$

$$\lambda_2 - \lambda_1 = 0$$

$$\lambda_1 \geq 0$$

$$\lambda_2 \geq 0$$

$$y^{(1)}(w_1 x_1^{(1)} + w_2 x_2^{(1)} + b) - 1 \geq 0$$

$$y^{(2)}(w_1 x_1^{(2)} + w_2 x_2^{(2)} + b) - 1 \geq 0$$

$$\lambda_1(y^{(1)}(w_1 x_1^{(1)} + w_2 x_2^{(1)} + b) - 1) = 0$$

$$\lambda_2(y^{(2)}(w_1 x_1^{(2)} + w_2 x_2^{(2)} + b) - 1) = 0$$

$$\frac{1}{2} - \lambda_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0$$

$$\lambda_2 - \lambda_1 = 0$$

$$\lambda_1 \geq 0$$

$$\lambda_2 \geq 0$$

$$2w_1 + w_2 + b - 1 \geq 0$$

$$-w_1 - w_2 - b + 1 \geq 0$$

$$\lambda_1(2w_1 + w_2 + b - 1) = 0$$

$$\lambda_2(-w_1 - w_2 - b + 1) = 0$$

3 Continuing the above problem. Write down the dual SVM optimization problem with linear kernel. (Hint: No need to derive the dual problem, but just plug the data in the dual problem in the lecture note).

$$\max_{\lambda} \quad L(\lambda) = \lambda_1 + \lambda_2 - \frac{5}{2}$$

$$\text{s.t.} \quad \lambda_1 - \lambda_2 = 0$$

$$\lambda_1,\ \lambda_2 \geq 0$$

4 Continue the above question. Given the dual parameters $\alpha = [0.5, 1]$, which may not be optimal, recover the primal parameter w (using the first-order condition of the KKT condition of SVM). Use linear kernel and solve for b by noting that any support vector on the margin should have $y^{(i)}(w^T x^{(i)} + b) = 1$. Then use the dual parameters to classify the test example $x = [1, 1]^T$ by computing the classifier output $h(x; \alpha)$.

$$w = \lambda_1 y^{(1)} x^{(1)} + \lambda_2 y^{(2)} x^{(2)}$$

$$= 0.5 \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$$

$$y^{(1)}(w_1 x_1^{(1)} + w_2 x_2^{(1)} + b) = 1$$

$$1.5 + b = 1$$

$$y^{(2)}(w_1 x_1^{(2)} + w_2 x_2^{(2)} + b) = 1$$

$$1.5 - b = 1$$

$$1.5 - b = 1.5 + b$$

$$b = 0$$

$$h(x; \lambda) = \sum_{i=1}^{m} \lambda_i y^{(i)} (x^{(i)})^T (x) + b$$

$$= 0.5[2, 1][1, 1]^T + [-1, 1][1, 1]^T$$

$$= 1.5$$

5 (Graduate only) Prove that there is a function $\psi([x_1, x_2])$, that maps from the input space $\mathbb{R}^2$ to a higher dimensional space $\mathbb{R}^n$, $n > 2$, so that the polynomial kernel $k(x, z) = (\langle x, z\rangle + 1)^2$ can be written as $\langle \psi(x), \psi(z)\rangle$.

Hints: refer to PRML Eq. (6.12) for help and write down the formula of the function ψ.

We are asked to prove that some function $\psi([x_1, x_2]$ maps from the $\mathbb{R}^2$ to a larger dimensional space such that polynomial kernal function $k(x, z) = (\langle x, z\rangle + 1)^2$ can be written as $\langle \psi(x), \psi(z)\rangle$.

We will start by multiplying the singular inner terms of the two dimensional column vectors.

$$
\begin{aligned}
k(x, z) &= (\langle x, z\rangle + 1)^2 \\
&= (x^T z + 1)^2 \\
&= (x_1 z_1 + X_2 z_2 + 1)^2 \\
&= x_1^2 z_1^2 + 2x_1 z_1 x_2 z_2 + 2x_1 z_1 + 2x_2 z_2 + x_2^2 z_2^2 + 1
\end{aligned}
$$

Thus we can see the mapping function ψ will have the following characteristics

$$
\begin{aligned}
&(x_1^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, x_2^2, 1)^T (z_1^2, \sqrt{2}z_1 z_2, \sqrt{2}z_1, \sqrt{2}z_2, z_2^2, 1) \\
&= \psi(x)^T \psi(z)
\end{aligned}
$$

We can easily see that we have mapped the column vector from one input space ($\mathbb{R}^2$) to a higher order dimensional space. The $\psi(x)$ function (in this two dimensional input space example) maps to ($\mathbb{R}^6$) space. This function performs operations on the inner values of the input vector and thus we can see that the inner product $\langle \psi(x), \psi(z)\rangle$ is the same as $k(x, z) = (\langle x, z\rangle + 1)^2$ and we have already shown that ψ maps to a higher dimensional space.

To solidify, given the two input space vector $x = [x_1, x_2]$. The function $\psi(x) = [x_1^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, x_2^2, 1]$.