Math for Ai final project

Ali Asheri nejad

February 13, 2024

We are looking to solve this equation in our primal problem.

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

subject to :

$$y_i(w.x_i + b) \geq 1 - \varepsilon \qquad and \qquad \varepsilon \geq 0$$

we can write min term and constraint in one equation :

$$L_p = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \varepsilon - \sum_{i=1}^{N} \lambda_i \Big(y_i(wx_i + b) - 1 + \varepsilon_i \Big)$$

We can find the optimal answer in each step by taking the derivative with respect to the variables in the equation and also determine the conditions . If we insert the obtained answers in the primal equation, we will reach the dual equation .

$$\frac{\partial L}{\partial w} = 0 \longrightarrow w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \longrightarrow w = \sum_{i=1}^{N} \lambda_i y_i x_i \tag{1}$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow \sum_{i=1}^{N} \lambda_i y_i = 0 \tag{2}$$

$$\frac{\partial L}{\partial \varepsilon} = 0 \longrightarrow c - \lambda_i - \mu_i = 0 \longrightarrow c = \lambda_i + \mu_i \tag{3}$$

$$\frac{\partial L}{\partial \lambda} = 0 \longrightarrow y_i(wx_i + b) - 1 + \varepsilon_i \geq 0 \longrightarrow y_i(wx_i + b) \geq 1 - \varepsilon_i \quad (4)$$

$$\frac{\partial L}{\partial \mu} = 0 \longrightarrow \varepsilon_i \geq 0 \quad (5)$$

$$\lambda > 0 \quad (6)$$

$$\mu_i > 0 \quad (7)$$

$$(3), (6), (7) \longrightarrow 0 < \lambda < c$$

So we can reach the dual equation by placing the above relationships in the primal equation (In order to satisfy the $\varepsilon \geq 0$ condition I can add the semester $\sum_{i=1}^{N} \mu_i \varepsilon_i$) :

$$\begin{aligned} L_p &= \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\varepsilon - \sum_{i=1}^{N}\lambda_i\Big(y_i(wx_i + b) - 1 + \varepsilon_i\Big) - \sum_{i=1}^{N}\mu_i\varepsilon_i \\ &= \frac{1}{2}W\sum_{i=1}^{N}\lambda_i y_i x_i + c\sum_{i=1}^{N}\varepsilon_i - \sum_{i=1}^{N}\lambda_i y_i w x_i - \sum_{i=1}^{N}\lambda_i y_i b + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i\varepsilon_i - \sum_{i=1}^{N}\mu_i\varepsilon_i \\ &= \frac{1}{2}W\sum_{i=1}^{N}\lambda_i y_i x_i + c\sum_{i=1}^{N}\varepsilon_i - w\sum_{i=1}^{N}\lambda_i y_i x_i - b\sum_{i=1}^{N}\lambda_i y_i + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i\varepsilon_i - \sum_{i=1}^{N}\mu_i\varepsilon_i \end{aligned}$$

We must apply the following conditions to above equation :

$$\sum_{i=1}^{N}\lambda_i y_i = 0 \longrightarrow -b\sum_{i=1}^{N}\lambda_i y_i = 0 \quad (8)$$

$$\frac{1}{2}W\sum_{i=1}^{N}\lambda_i y_i x_i - w\sum_{i=1}^{N}\lambda_i y_i x_i = -\frac{1}{2}W\sum_{i=1}^{N}\lambda_i y_i x_i \quad (9)$$

$$c = \sum_{i=1}^{N} \lambda_i + \sum_{i=1}^{N} \mu_i \longrightarrow c \sum_{i=1}^{N} \varepsilon_i = \sum_{i=1}^{N} \lambda_i \varepsilon_i + \sum_{i=1}^{N} \mu_i \varepsilon_i \tag{10}$$

so we can write equation :

$$\begin{aligned}
&= -\frac{1}{2} W \sum_{i=1}^{N} \lambda_i y_i x_i + \sum_{i=1}^{N} \lambda_i \\
&= -\frac{1}{2} (\sum_{i=1}^{N} \lambda_i y_i x_i)(\sum_{i=1}^{N} \lambda_i y_i x_i) + \sum_{i=1}^{N} \lambda_i \\
&= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i x_j + \sum_{i=1}^{N} \lambda_i
\end{aligned}$$

Dual problem is $-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i x_j + \sum_{i=1}^{N} \lambda_i$ subject to :

$0 < \lambda_i \leq c$ and $\sum_{i=1}^{N} \lambda_i y_i = 0$

How we can solve Primal problem with Dual problem?

we can get λ_i from solve Dual problem $-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i x_j + \sum_{i=1}^{N} \lambda_i$. then get W from this condition : $w = \sum_{i=1}^{N} \lambda_i y_i x_i$

Usually solving each of these problems will have the same result as we say in this case we have `Strong Duality` .

Code :

1-primal :

we want $min \quad w_1^2 + w_2^2 + w_3^2 + w_4^2 + C\sum_{i=1}^{N}\varepsilon_i \qquad w_1, w_2, w_3, w_4$ are weights of Axis and b is bias and ε_i are slack points .To solve the primal problem , we must also consider the slack points . So we have to consider slack points according to the number of samples we have . so we can write :

$$w_1^2 + w_2^2 + w_3^2 + w_4^2 + C\sum_{i=1}^{N}\varepsilon_i =$$

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 & b & \varepsilon_1 & .... & \varepsilon_n \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ b \\ \varepsilon_1 \\ ... \\ \varepsilon_n \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & C_1 & C_2 & ... & C_n & .... \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ b \\ \varepsilon_1 \\ ... \\ \varepsilon_n \end{bmatrix}$$

now we should apply condition to our problem :
$y_i(w.x_i + b) \geq 1 - \varepsilon_i \longrightarrow -y_i(w.x_i + b) + \varepsilon_i \leq -1$

and other condition is : $\varepsilon_i \geq 0$

We must apply both conditions in the same matrix :

$$\begin{bmatrix} -y_1x_{11} & -y_1x_{12} & -y_1x_{13} & -y_1x_{14} & -y_1 & -1 & 0 & ... & 0 \\ -y_2x_{11} & -y_2x_{12} & -y_2x_{13} & -y_2x_{14} & -y_2 & 0 & -1 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... & ... \\ -y_nx_{11} & -y_nx_{12} & -y_nx_{13} & -y_nx_{14} & -y_n & 0 & 0 & ... & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ b \\ \varepsilon_1 \\ ... \\ \varepsilon_n \end{bmatrix}$$

$$\leq \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ ... \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix}$$

2-Dual :

we want find λ from maximize $-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i x_j + \sum_{i=1}^{N}\lambda_i$

As is clear from the formula, this problem is also Quadratic

in matrix form :

1.matrix =

$$A = -\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ ... \\ y_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & ... & y_n \end{bmatrix} * \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & ... & x_{n1} \\ x_{12} & x_{22} & ... & x_{n2} \\ x_{13} & x_{23} & ... & x_{n3} \\ x_{1,4} & x_{2,4} & ... & x_{n,4} \end{bmatrix}$$

$$\begin{bmatrix} \lambda & \lambda_2 & \lambda_3 & \lambda_4 & ... & \lambda_n \end{bmatrix} A \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ ... \\ \lambda_n \end{bmatrix} + \begin{bmatrix} \lambda & \lambda_2 & \lambda_3 & \lambda_4 & ... & \lambda_n \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ ... \\ 1 \end{bmatrix}$$

2.matrix =

we should show $0 < \lambda_i \leq c$ in matrix form .
we can write $-\lambda_i < 0$ and $\lambda_i < c$. so we can write in matrix form :

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 & ... & \lambda_{2n} \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 & ... & 0 \\ 0 & -1 & 0 & ... & 0 \\ 0 & 0 & -1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -1 \\ 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \\ c \\ c \\ c \\ ... \\ c \end{bmatrix}$$

3.matrix

we want to show $\sum_{i=1}^{N} \lambda_i y_i = 0$ in matrix form so we can write :

$$\begin{bmatrix} y_1 & y_2 & y_3 & ... & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ ... \\ \lambda_n \end{bmatrix} = 0$$

After defining the matrices and obtaining the λ_i, we can obtain the weights and bias through the conditions : $w = \sum_{i=1}^{N} \lambda_i y_i x_i$

matrix form :

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ ... \\ \lambda_n \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ ... \\ y_n \end{bmatrix} * \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \\ ... & ... & ... & ... \\ w_{n1} & w_{n2} & w_{n3} & w_{n4} \end{bmatrix}$$

$$\text{W} = \begin{bmatrix} \sum_{i=1}^{N} w_{i1} & \sum_{i=1}^{N} w_{i2} & \sum_{i=1}^{N} w_{i3} & \sum_{i=1}^{N} w_{i4} \end{bmatrix}$$

also for get bias we should use from $\lambda_i(y_i(wx_i + b)) = 1$:

$\lambda_i(y_i(wx_i + b)) = 1 \longrightarrow \lambda_1(y_1(wx_1 + b)) + \lambda_2(y_2(wx_2 + b)) + ... + \lambda_n(y_n(wx_n + b))$

$$y_i w x_i + y_i b = 1$$

$$y_i b = 1 - y_i w x_i$$

$$b = \frac{1}{y} - wx_i \ , \ \frac{1}{y} = y$$

$$b = y - wx_i$$