Question 1

(1)

Given that

1. The training data: $T = {(x_1,y_1), ..., (x_n,y_n)}$, where $x_i \in R^n, y_i \in {-1, 1}, i = 1, 2, ..., N$
2. The hyperplane: $wx+b = 0$

The margin between a sample (x_i, y_i) and the hyperplane is defined accordingly,

$$ \gamma_i = y_i(\frac{w}{||w||}x_i + \frac{b}{||w||}) $$

The minimal distance between the hyperplane and the training data is

$$ \gamma = \underset{i=1,2,...,N}{min} \gamma_i $$

The problem is equivalent to the below optimization problem

$$ \begin{equation} \begin{aligned} \underset{w,b}{max} &\ \gamma \ s.t. &\ y_i(\frac{w}{||w||}x_i + \frac{b}{||w||}) \geq r, i =1,2,...,N \end{aligned} \end{equation} $$

Define $\gamma = \frac{\gamma'}{||w||}$, the equation (1) can be rewritten as

$$ \begin{equation} \begin{aligned} \underset{w,b}{max} &\ \frac{\gamma'}{||w||} \ s.t. &\ y_i(wx_i+b) \geq \gamma', i =1,2,...,N \end{aligned} \end{equation} $$

which is equivalent to the optimization problem (3)

$$ \begin{equation} \begin{aligned} \underset{w,b}{min} &\ \frac{1}{2}||w||^2 \ s.t. &\ y_i(wx_i+b) - 1 \geq 0, i =1,2,...,N \end{aligned} \end{equation} $$

Rewrite the equation (3) to the Lagrange multiplier format

$$ L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^N \alpha_i y_i (w\cdot x_i + b) + \sum_{i=1}^N \alpha_i $$

we have

$$ \underset{w,b}{min} \ \underset{\alpha \geq 0}{max} \ L(w, b, \alpha) = \underset{\alpha \geq 0}{max} \ \underset{w,b}{min} \ L(w, b, \alpha) $$

$$ \begin{aligned} \underset{w,b}{min} \ L(w, b, \alpha) &=\underset{w,b}{min} \ \frac{1}{2}||w||^2 - \sum_{i=1}^N \alpha_i y_i (w\cdot x_i + b) + \sum_{i=1}^N \alpha_i \ \end{aligned} $$

Let,

$$ \begin{aligned} \frac{\partial L(w, b, \alpha)} {\partial w} &= w - \sum_{i=1}^N \alpha_iy_ix_i = 0 \ \frac{\partial L(w, b, \alpha)} {\partial b} &= - \sum_{i=1}^N \alpha_iy_i = 0 \end{aligned} $$

we have

$$ \begin{equation} w = \sum_{i=1}^N \alpha_iy_ix_i ,\quad \sum_{i=1}^N \alpha_iy_i = 0 \end{equation} $$

Substituting equations (4) to $\underset{w,b}{min} \ L(w, b, \alpha)$, we have

$$ \begin{equation} \begin{aligned} \underset{w,b}{min} \ L(w, b, \alpha) &=\underset{w,b}{min} \ \frac{1}{2}||\sum_{i=1}^N \alpha_iy_ix_i||^2 - \sum_{i=1}^N \alpha_i y_i (\sum_{i=1}^N \alpha_iy_ix_i\cdot x_i + b) + \sum_{i=1}^N \alpha_i \ &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_jy_iy_j(x_i\cdot x_j) - \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_jy_iy_j(x_i\cdot x_j) -\ &\quad \sum_{i=1}^N \alpha_iy_ib + \sum_{i=1}^N \alpha_iy_i \ & = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_jy_iy_j(x_i\cdot x_j) + \sum_{i=1}^N \alpha_iy_i \end{aligned} \end{equation} $$

Due to the weak dual theorem, the primal optimization problem

$$ \begin{equation} \begin{aligned} \underset{w,b}{min} &\ \frac{1}{2}||w||^2 \ s.t. &\ y_i(wx_i+b) - 1 \geq 0, i =1,2,...,N \end{aligned} \end{equation} $$

is weak dual to the dual problem

$$ \begin{equation} \begin{aligned} \underset{\alpha}{max} &\ -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_jy_iy_j(x_i\cdot x_j) + \sum_{i=1}^N \alpha_iy_i \ s.t. &\ \sum_{i=1}^N \alpha_iy_i = 0, i =1,2,...,N \end{aligned} \end{equation} $$

If there is $(w^, b^, \alpha^)$ and (w^, b^) is a feasible solution to the primal and $(|alpha^)$ is feasible to the dual, then $(w^, b^, \alpha^*)$ is the optimal solution.

(2)

If we plot the 'support vectors', they are hyperplace. (In 2D, they are lines while they are planes in 3D)