Machine Learning Exercise 4

何舜成

2012011515

1. Give a proof of $VC(\mathcal{H}) = VC(\Phi)$

<u>Proof</u>:

Given the condition $VC(\Phi) = d$, we know $\exists z_1, z_2, \cdots, z_d$ such that

$$|\{(\phi(z_1), \cdots, \phi(z_d))|\phi \in \Phi\}| = 2^d$$

and there are no $z_1, \cdots, z_d, z_{d+1}$ such that

$$|\{(\phi(z_1), \cdots, \phi(z_{d+1})|\phi \in \Phi\}| = 2^{d+1}$$

Define set $\mathcal{S}_Z = \{(f(x_1), \cdots, f(x_d))|f \in \mathcal{H}\}$ and $\mathcal{T}_Z = \{(\phi(z_1), \cdots, \phi(z_d))|\phi \in \Phi\}$ for a fixed sequence $Z = (z_1, \cdots, z_d), z_i = (x_i, y_i)$.

Let $g_Z(S) = (I(y_1 \neq s_1), \cdots, I(y_d \neq s_d)) = (y_1, \cdots, y_d) \oplus (s_1, \cdots, s_d)$ ($\oplus$ stands for XOR). Since for all binary sequence $S_1, S_2, S_1 \neq S_2$, $g_Z(S_1) \neq g_Z(S_2)$ holds true, and for all binary sequence T, $\exists S_T$ such that $g_Z(S_T) = T$, function $g_Z(\cdot) : \mathcal{S}_Z \to \mathcal{T}_Z$ is **bijection**, and two sets $\mathcal{S}_Z$ and $\mathcal{T}_Z$ are equipotent.

If $VC(\Phi) = d$, then $\exists z_1, z_2, \cdots, z_d$ such that

$$|\{(f(x_1), \cdots, f(x_d))|f \in \mathcal{H}\}| = |\{(\phi(z_1), \cdots, \phi(z_d))|\phi \in \Phi\}| = 2^d$$

and for all $z_1, \cdots, z_d, z_{d+1}$

$$|\{(f(x_1), \cdots, f(x_{d+1}))|f \in \mathcal{H}\}| = |\{(\phi(z_1), \cdots, \phi(z_{d+1})|\phi \in \Phi\}| \neq 2^{d+1}$$

Therefore $VC(\mathcal{H}) = d$, and vice versa.

2. Prove the equivalence of the two optimization problems

$$\begin{aligned} \max_{w,b,t} \quad & t \\ s.t. \quad & y_i(w^T x_i + b) \geq t, \forall i \in [n] \\ & ||w|| = 1 \end{aligned}$$

and

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ s.t. \quad & y_i(w^T x_i + b) \geq 1, \forall i \in [n] \end{aligned}$$

Proof:
The primal problem $\Leftrightarrow$

$$\begin{aligned} \max_{w,b,t} \quad & \frac{t}{||w||} \\ s.t. \quad & y_i((\frac{w}{t})^T x_i + \frac{b}{t}) \geq 1, \forall i \in [n] \\ & ||w|| = 1 \end{aligned}$$

Let $v = \frac{w}{t}$ and $c = \frac{b}{t}$, then the primal $\Leftrightarrow$

$$\begin{aligned} \max_{v,c} \quad & \frac{1}{||v||} \\ s.t. \quad & y_i(v^T x_i + c) \geq 1, \forall i \in [n] \end{aligned}$$

It is obvious that maximizing $\frac{1}{||v||}$ is equivalent to minimizing $\frac{1}{2}||v||^2$, and the primal $\Leftrightarrow$

$$\begin{aligned} \min_{v,c} \quad & \frac{1}{2}||v||^2 \\ s.t. \quad & y_i(v^T x_i + c) \geq 1, \forall i \in [n] \end{aligned}$$

3. Give the dual problem of the latter one of Ex 2, namely the dual problem of

$$\begin{array}{ll} \min_{w,b} & \frac{1}{2}||w||^2 \\ s.t. & y_i(w^T x_i + b) \geq 1, \forall i \in [n] \end{array}$$

Solution:

Define the Langrangian function

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + b) - 1), \forall i \in [n], \alpha_i \geq 0$$

Set the partial derivatives to zero, w.r.t. w and b respectively, and we get

$$\begin{array}{lll} \frac{\partial \mathcal{L}}{\partial w} = 0 & \Rightarrow & w = \sum_{i=1}^{n} \alpha_i y_i x_i \\ \frac{\partial \mathcal{L}}{\partial b} = 0 & \Rightarrow & \sum_{i=1}^{n} \alpha_i y_i = 0 \end{array}$$

Therefore

$$\begin{aligned} \mathcal{L}(w, b, \alpha) &= \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + b) - 1) \\ &= \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i y_i w^T x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i \\ &= \frac{1}{2} w^T \sum_{i=1}^{n} \alpha_i y_i x_i - w^T \sum_{i=1}^{n} \alpha_i y_i x_i - b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \\ &= -\frac{1}{2} w^T \sum_{i=1}^{n} \alpha_i y_i x_i + \sum_{i=1}^{n} \alpha_i \\ &= -\frac{1}{2} (\sum_{i=1}^{n} \alpha_i y_i x_i)^T \sum_{i=1}^{n} \alpha_i y_i x_i + \sum_{i=1}^{n} \alpha_i \\ &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \end{aligned}$$

The dual problem can be described as follow

$$
\begin{aligned}
\max_\alpha \quad & \sum_{i=1}^{n} - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\
s.t. \quad & \alpha_i \geq 0, \forall i \in [n] \\
& \sum_{i=1}^{n} \alpha_i y_i = 0
\end{aligned}
$$