CES417T - Homework 1

467261 - Yifu Wang

2018 - 09 - 07

1 (LFD Problem 1.3)

Follow the instruction step by step:

(a) Since $\mathbf{w}^*$ is a separation of the data. We have $y_n = \text{sign}(\mathbf{w}^{*T}\mathbf{x}_n) \neq 0$, thus $\rho = \min_{[1,N]} y_n(\mathbf{w}^{*T}\mathbf{x}_n) > 0$.

(b) By the definition of ρ and $\mathbf{w}(t+1) = \mathbf{w}(t) + y(t)\mathbf{x}(t)$, we have

$$\begin{aligned}\left(\mathbf{w}^T(t+1) - \mathbf{w}^T(t)\right)\mathbf{w}^* &= (\mathbf{w}(t+1) - \mathbf{w}(t))^T\mathbf{w}^* \\ &= \left(y(t)\mathbf{x}(t)\right)^T\mathbf{w}^* \\ &= y(t)\mathbf{x}^T(t)\mathbf{w}^* \\ &= y(t)\left(\mathbf{w}^*\mathbf{x}^T(t)\right) \\ &\geq \rho\end{aligned}$$

Denote $\rho_t = \left(\mathbf{w}^T(t+1) - \mathbf{w}^T(t)\right)\mathbf{w}^*$. Then

$$\begin{aligned}\mathbf{w}^T(t)\mathbf{w}^* &= \mathbf{w}^T(0)\mathbf{w}^* + \sum_{k=0}^{t-1}\rho_k \\ &= \sum_{k=0}^{t-1}\rho_k \\ &\geq t\rho\end{aligned}$$

(c) Since $y(t)\mathbf{w}^T(t)\mathbf{x}(t) \leq 0$, we have

$$\begin{aligned}\|\mathbf{w}(t)\|^2 - \|\mathbf{w}(t-1)\|^2 - \|\mathbf{x}(t-1)\|^2 &= \mathbf{w}^T(t)\mathbf{w}(t) - \mathbf{w}^T(t-1)\mathbf{w}(t-1) - x^T(t-1)x(t-1) \\ &= (\mathbf{w}(t) - \mathbf{w}(t-1))^T(\mathbf{w}(t) + \mathbf{w}(t-1)) - x^T(t-1)x(t-1) \\ &= \left(y(t-1)\mathbf{x}(t-1)\right)^T(\mathbf{w}(t) + \mathbf{w}(t-1)) - x^T(t-1)x(t-1) \\ &\leq y(t-1)\mathbf{x}^T(t-1)\mathbf{w}(t) - x^T(t-1)x(t-1) \\ &= y(t-1)\mathbf{x}^T(t-1)\left(\mathbf{w}(t) - y(t-1)\mathbf{x}(t-1)\right) \\ &= y(t-1)\mathbf{x}^T(t-1)\mathbf{w}(t-1) \\ &\leq 0\end{aligned}$$

(d) Based on **(c)**, we have

$$\begin{aligned}\|\mathbf{w}(t)\|^2 &= \sum_{k=1}^{t}\left(\|\mathbf{w}(k)\|^2 - \|\mathbf{w}(k-1)\|^2\right) \\ &\leq \sum_{k=1}^{t}\|\mathbf{x}(t-1)\|^2 \\ &\leq tR^2\end{aligned}$$

namely, $\|\mathbf{w}(t)\| \leq \sqrt{t}R$.

(e) Based on **(b)** and **(d)**, we have

$$\frac{\mathbf{w}^T(t)\mathbf{w}^*}{\|\mathbf{w}(t)\|} \ge \frac{t\rho}{\sqrt{t}R} = \sqrt{t}\frac{\rho}{R}$$

Hence $t \le \frac{R^2\|\mathbf{w}^*\|^2}{\rho^2}$, because $\frac{\mathbf{w}(t)}{\|\mathbf{w}(t)\|} = 1$.

Finally, we get an upper bound of t, which means PLA will converge.