2.3 感知器算法的收敛性（*Guarantee of PLA*）

课后证明：

(1)：针对线性可分训练样本集，*PLA*算法中，当 $\boldsymbol{w}_0 = \boldsymbol{0}$，在对分错样本进行了*T*次修正后，下式成立：$\frac{\mathbf{w}_f^T}{\|\mathbf{w}_f\|}\frac{\mathbf{w}_T}{\|\mathbf{w}_T\|} \ge \sqrt{T}\cdot constant$

(2) ，针对线性可分训练样本集，*PLA*算法中，假设对分错样本进行了*T*次修正后得到的分类面不再出现错分状况，定义：

$R^2 = \max_n \|\mathbf{x}_n\|^2$，$\rho = \min_n y_n \frac{\mathbf{w}_f^T}{\|\mathbf{w}_f\|}\mathbf{x}_n$，证明：$\mathrm{T} \le \frac{R^2}{\rho^2}$

证明(1) $w_f^T w_T = w_f^T(w_{T-1} + y_{n(t)} x_{n(t)}) = w_f^T w_{T-1} + w_f^T y_{n(t)} x_{n(t)}$

令 $w_f^T y_{n(t)} x_{n(t)} \ge \min_n y_n w_f^T x_n = M$

$\therefore w_f^T w_T \ge w_f^T w_{T-1} + M \ge w_f^T w_{T-1} + 2M \ge \cdots \ge TM$

$\because \|w_T\|^2 = \|w_{T-1} + y_{n(t)} x_{n(t)}\| = \|w_{T-1}\|^2 + 2 y_{n(t)} w_{T-1} x_{n(t)} + \|x_{n(t)}\|^2$

又∵该样本被分类错误而更新

$\therefore y_{n(t)} w_t^T x_{n(t)} \le 0$

$\therefore \|w_T\|^2 \le \|w_{T-1}\|^2 + \|x_{n(t)}\|^2 \le \|w_{T-2}\|^2 + 2\cdot\|x_{n(t)}\|^2 \le T\cdot\|x_{n(t)}\|^2$

令 $\|x_{n(t)}\|^2 \le \max_n \|x_n\|^2 = P^2$

$\therefore \|w_T\| \le \sqrt{T} P$

又∵柯西不等式 $w_f^T \cdot w_T \le \|w_T\|\|w_f\|$

$\therefore \frac{w_f w_T}{\|w_T\|\|w_f\|} \ge \sqrt{T}\frac{M}{P\|w_f\|} = \sqrt{T}\cdot constant$

(2) 由(1)得

$\frac{w_f^T w_T}{\|w_f\|} \ge T\rho$

$\|w_T\|^2 \le TR^2$

且 $w_f^T \cdot w_T \le \|w_f\|\|w_T\|$

$\therefore \frac{w_f^T w_T}{\|w_f\|\|w_T\|} = \sqrt{T}\frac{\rho}{R} \le 1$

$\therefore T \le \frac{R^2}{\rho^2}$