$$\begin{aligned}\boldsymbol{W}_f^T\boldsymbol{W}_{t+1} &= \boldsymbol{W}_f^T(\boldsymbol{W}_t + y_n(t)\boldsymbol{X}_n(t)) \\ &\geqslant \boldsymbol{W}_f^T\boldsymbol{W}_t + \min_n y_n(t)\boldsymbol{W}_f^T\boldsymbol{X}_n(t)\end{aligned}$$

且有$W_0 = 0$，故有$\boldsymbol{W}_f^T\boldsymbol{W}_T \geqslant T \cdot \min_n y_n \boldsymbol{W}_f^T\boldsymbol{X}_{\boldsymbol{n}}$；

又由于

$$\begin{aligned}\|\boldsymbol{W}_{t+1}\|^2 &= \|\boldsymbol{W}_t + y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &= \|\boldsymbol{W}_t\|^2 + 2y_n(t)\boldsymbol{W}_t^T\boldsymbol{X}_n(t) + \|y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &\leqslant \|\boldsymbol{W}_t\|^2 + 0 + \|y_n(t)\boldsymbol{X}_n(t)\|^2 \\ &\leqslant \|\boldsymbol{W}_t\|^2 + \max_n \|\boldsymbol{X}_n(t)\|^2\end{aligned}$$

故有$\|\boldsymbol{W}_T\| \leqslant \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}$；

综上所述，有

$$\begin{aligned}\frac{\boldsymbol{W}_f^T\boldsymbol{W}_T}{\|\boldsymbol{W}_f\|\|\boldsymbol{W}_T\|} &\geqslant \frac{T \cdot \min_n y_n \boldsymbol{W}_f^T\boldsymbol{X}_{\boldsymbol{n}}}{\|\boldsymbol{W}_f\| \cdot \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}} \\ &= \sqrt{T} \cdot constant\end{aligned}$$

4，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 次纠正后得到的分类面不再出现错分状况，定义：$\boldsymbol{R^2 = \max_n \|x_n\|^2}$ ，$\rho = \min_n y_n \frac{\mathbf{w}_f^T}{\|\mathbf{w}_f\|}\mathbf{x}_n$，试证明：$\mathrm{T} \leq \frac{\boldsymbol{R^2}}{\rho^2}$

证明：

$$\begin{aligned}\frac{\boldsymbol{W}_f^T\boldsymbol{W}_T}{\|\boldsymbol{W}_f\|\|\boldsymbol{W}_T\|} &\geqslant \frac{T \cdot \min_n y_n \boldsymbol{W}_f^T\boldsymbol{X}_{\boldsymbol{n}}}{\|\boldsymbol{W}_f\| \cdot \sqrt{T \cdot \max_n \|\boldsymbol{X}_n\|^2}} \\ &= \sqrt{T} \cdot \frac{\rho}{R}\end{aligned}$$