////////////////////// Bonus Problem ////////////////////////

21. We want to show that $y_{n(t)} w_{t+1}^T x_{n(t)} > 0$ with the new update rule. So, we substitute the w_{t+1} with the new one.

$$y_{n(t)}\mathbf{w}_{t+1}^T\mathbf{x}_{n(t)} = y_{n(t)}\{\mathbf{w}_t + y_{n(t)}\mathbf{x}_{n(t)}\left\lfloor \frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2} + 1 \right\rfloor\}^T \mathbf{x}_{n(t)}$$

$$= y_{n(t)}\{\mathbf{w}_t + \left\lfloor \frac{\overset{=-1}{-y_{n(t)}{}^2}\, \mathbf{w}_t^T \|\mathbf{x}_{n(t)}\|^2}{\|\mathbf{x}_{n(t)}\|^2} \right\rfloor + y_{n(t)}\mathbf{x}_{n(t)}\}\mathbf{x}_{n(t)}$$

$$= y_{n(t)}\mathbf{w}^T{}_t\mathbf{x}_{n(t)} - \left\lfloor y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} \right\rfloor + \underbrace{y_{n(t)}{}^2 \|\mathbf{x}_{n(t)}\|^2}_{=\|\mathbf{x}_{n(t)}\|^2} > 0$$

22. We want to prove that the PLA will halt in a perfect line if date is linear separable and using the new update rule.

If data is linear separable, it means there exist a $\boldsymbol{w_f}$ that every $\boldsymbol{x_n}$ correctly away from the line. Hence, we know that $y_{n(t)}\mathbf{w}_f^T\mathbf{x}_{n(t)} \geq \min_n y_n \mathbf{w}_f^T\mathbf{x}_n > 0$. Moreover, the update only occur when mistake; so, $y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} \leq 0$.

Now, we using the skill taught in class in lecture 2, in page 14-15. We want to prove that each update can make the $\boldsymbol{w_t}$ more close to the $\boldsymbol{w_f}$ with the new update rule.

$$\because y_{n(t)} \neq sign(\mathbf{w}_t^T\mathbf{x}_{n(t)}) \;\; \therefore \text{-}y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} > 0 \Rightarrow \left\lfloor \frac{\text{-}y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2} \right\rfloor \geq 0 \;\Rightarrow \left\lfloor \frac{\text{-}y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2} + 1 \right\rfloor \;\geq\; 1$$

First, we want to prove that first equation $\mathbf{w}_f^T\mathbf{w}_{t+1} \geq \mathbf{w}_f^T\mathbf{w}_t$ still hold.

$$\mathbf{w}_f^T\mathbf{w}_{t+1} = \mathbf{w}_f^T(\mathbf{w}_t + y_{n(t)}\mathbf{x}_{n(t)}\left\lfloor \frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2} + 1 \right\rfloor)$$

$$\geq \mathbf{w}_f^T(\mathbf{w}_t + y_{n(t)}\mathbf{x}_{n(t)})$$

$$\geq \mathbf{w}_f^T\mathbf{w}_t + \min_n y_{n(t)}\mathbf{w}_f^T\mathbf{x}_{n(t)}$$

$$\geq \mathbf{w}_f^T\mathbf{w}_t + 0$$

Second, we want to prove that the growth of $\|\mathbf{w}_{t+1}\|^2$ is limited.

$$
\begin{aligned}
\|\mathbf{w}_{t+1}\|^2 &= \left\|\mathbf{w}_t + y_{n(t)}\mathbf{x}_{n(t)}\left\lfloor \frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2}+1\right\rfloor\right\|^2 \\
&= \|\mathbf{w}_t\|^2 + \underbrace{2y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}_{\le 0}\left\lfloor \underbrace{\frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2}+1}_{\ge 1}\right\rfloor + \left\|y_{n(t)}\mathbf{x}_{n(t)}\left\lfloor \frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2}+1\right\rfloor\right\|^2 \\
&\le \|\mathbf{w}_t\|^2 + 2y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} + \left\|y_{n(t)}\mathbf{x}_{n(t)}\left\lfloor \frac{-y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)}}{\|\mathbf{x}_{n(t)}\|^2}+1\right\rfloor\right\|^2 \\
&\le \|\mathbf{w}_t\|^2 + 2y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} + \left\|\frac{-y_{n(t)}{}^2\mathbf{w}_t^T\cancel{\mathbf{x}_{n(t)}{}^2}}{\cancel{\|\mathbf{x}_{n(t)}\|^2}} + y_{n(t)}\mathbf{x}_{n(t)}\right\|^2 \\
&= \|\mathbf{w}_t\|^2 + 2y_{n(t)}\mathbf{w}_t^T\mathbf{x}_{n(t)} + \|-\mathbf{w}_t + y_{n(t)}\mathbf{x}_{n(t)}\|^2 \\
&= 2\|\mathbf{w}_t\|^2 + \|\mathbf{x}_{n(t)}\|^2 \\
&\le 2\|\mathbf{w}_t\|^2 + \max_n\|\mathbf{x}_{n(t)}\|^2
\end{aligned}
$$

From the above result, we can derive that if we start from $\mathbf{w}=0$, $\frac{\mathbf{w}_f^T\mathbf{w}_t}{\|\mathbf{w}_f\|\|\mathbf{w}_t\|} \ge f(t)\times const.$

$$
\begin{aligned}
\frac{\mathbf{w}_f^T \mathbf{w}_t}{\|\mathbf{w}_f\|\|\mathbf{w}_t\|} &\geq \frac{\mathbf{w}_f^T \times t \times \min_n y_{n(t)} \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|\|\mathbf{w}_t\|} \\
&= t \times \frac{\min_n y_{n(t)} \mathbf{w}_f^T \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|} \times \frac{1}{\sqrt{\|\mathbf{w}_t\|^2}} \\
&\geq t \times \frac{\min_n y_{n(t)} \mathbf{w}_f^T \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|} \times \frac{1}{\sqrt{2\|\mathbf{w}_t\|^2 + \max_n \|\mathbf{x}_{n(t)}\|^2}} \\
&\geq t \times \frac{\min_n y_{n(t)} \mathbf{w}_f^T \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|} \times \frac{1}{\sqrt{2t \max_n \|\mathbf{x}_{n(t)}\|^2}} \\
&\geq t \times \frac{\min_n y_{n(t)} \mathbf{w}_f^T \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|} \times \frac{1}{\sqrt{t}\sqrt{2 \max_n \|\mathbf{x}_{n(t)}\|^2}} \\
&= \frac{t}{\sqrt{t}} \times \frac{\min_n y_{n(t)} \mathbf{w}_f^T \mathbf{x}_{n(t)}}{\|\mathbf{w}_f\|} \times \frac{1}{\sqrt{2 \max_n \|\mathbf{x}_{n(t)}\|^2}} \\
&\geq \sqrt{t} \times const.
\end{aligned}
$$

Finally, we conclude that $\mathbf{w}_t$ can close to the $\mathbf{w}_f$ by each update step with new update rule and maximum value of $\frac{\mathbf{w}_f^T \mathbf{w}_t}{\|\mathbf{w}_f\|\|\mathbf{w}_t\|}$ will be 1.