Ex 4.6

(a) From 4.41 we know that

① if $y_i = 1$ is misclassified then $x_i^T \beta_1 + \beta_0 < 0$

② if $y_i = -1$ is misclassified then $x_i^T \beta_1 + \beta_0 > 0$

So if i is correctly classified then $y_i \beta^T x_i^* > 0$ (✱)

Then if the classes are seperable and all observations are correctly classified then (✱) is the case $\forall i$.
Furthermore, there exists some minimum value $K \in 1, \ldots, N$ that minimises (✱) giving $y_K \beta^T x_K^* = K$. (K = Scalar)

Thus we can rewrite (✱) as:

$$y_i \beta^T x_i^* \geq K \qquad \forall i$$

$$\therefore \quad y_i \beta_{sep}^T x_i^* \geq 1 \qquad \forall i \qquad \text{where } \beta_{sep}^T = \beta^T / K$$

(b)

$$\| \beta_{New} - \beta_{sep} \|^2 = \| \beta_{old} + y_i Z_i - \beta_{sep} \|^2$$

①

$$= \| \beta_{old} \|^2 + \| y_i Z_i \|^2 - \| \beta_{sep} \|^2 + 2 \beta_{old}^T y_i Z_i - 2 \beta_{old}^T \beta_{sep} - 2 y_i \beta_{sep}^T Z_i$$

aside//

$$y_i \beta_{sep}^T Z_i \geq 1 \qquad \text{(From Part (a))}$$

$$y_i \beta_{old}^T Z_i < 0 \qquad \text{(Since it was misclassified and using (✱))}$$

$$\Rightarrow -y_i \beta_{old}^T Z_i > 0$$

Subbing these facts into ① we obtain

$$① \leq \| \beta_{old} \|^2 + \| \beta_{sep} \|^2 - 2 \beta_{old}^T \beta_{sep} + \| y_i Z_i \|^2 - 2(0 + 1)$$

$$= \| \beta_{old} - \beta_{sep} \|^2 + \| y_i Z_i \|^2 - 2 \qquad ②$$

Now $\| y_i Z_i \|^2 = |y_i| \, \| Z_i \|^2$

$|y_i| = 1$ and $\| Z_i \|^2 = \left\| \frac{x_i^*}{\|x_i\|^*} \right\|^2 = 1$

$\Rightarrow$ ② $= \| \beta_{old} - \beta_{sep} \|^2 + 1 - 2$

Overall:

$$\| \beta_{New} - \beta_{sep} \|^2 \leq \| \beta_{old} - \beta_{sep} \|^2 - 1$$ as required.

Therefore β_{sep} can be found in at most $\| \beta_{start} - \beta_{sep} \|^2$ steps.