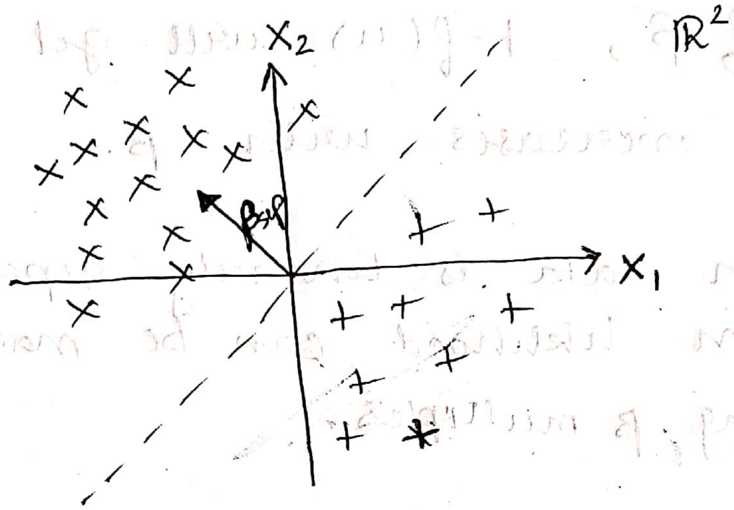
2.

a) $$L_\beta(X,y) = \prod_{i=1}^{n} L_\beta(x_i, y_i)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{1+\exp(-x_i^T\beta)}\right)^{y_i} \left(\frac{\exp(-x_i^T\beta)}{1+\exp(-x_i^T\beta)}\right)^{1-y_i} \quad -①$$

b)



c) Given the property $f(u) = \frac{1}{1+\exp(-u)}$ being strictly increasing, lets take a look at the first part of eq ①

$\therefore \quad u = x_i^T\beta$

When the data is linearly seperable, $x_i^T\beta$ will be positive for all training points with $y_i = 1$.

So this part can always be made larger by increasing β (by multiples) - (deduced from that fact that $f(u)$ is strictly increasing

The second term in ①, is

$$\left(\frac{\exp(-x_i^T\beta)}{1+\exp(-x_i^T\beta)}\right)^{1-y_i}$$

which is equal to $1-f(u)$. Since the data is linearly seperable, $x_i^T\beta < 0$ for all training items with $y_i = 0$, on increasing multiples of β, $1-f(u)$ will get larger.

This part also increases with β.

Hence, when data is linearly seperable, the Maximum likelihood can be made larger by increasing β multiples.