Question2

2a.

$$\begin{aligned} y_i(w^T x_i + b) &\geq 1 - \xi_i \\ \xi_i &\geq 1 - y_i(w^T x_i + b) \end{aligned} \tag{1}$$

If $1 - y_i(w^T x_i + b) > 0$, $\xi_i^2 \geq (1 - y_i(w^T x_i + b))^2$;
if $1 - y_i(w^T x_i + b) < 0$, $|\xi_i| > 0 > 1 - y_i(w^T x_i + b)$.

$$|\xi_i|^2 \geq \begin{cases} (1 - y_i(w^T x_i + b))^2, & y_i(w^T x_i + b) < 1 \\ 0, & y_i(w^T x_i + b) > 1 \end{cases} \tag{2}$$

$$\|\xi\|^2 = \sum_i |\xi_i|^2$$

Substitute above into $G(w, b, \xi)$, we have

$$\begin{aligned} G(w, b, \xi) &= \|w\|^2 + \alpha\|\xi\|^2 \\ &= \|w\|^2 + \alpha \sum_i |\xi_i|^2 \\ &= \|w\|^2 + \alpha \sum_i \max(0, (1 - y_i(w^T x_i + b)))^2 \\ &= F(w, b) \end{aligned} \tag{3}$$

2b.

We construct the Lagrangian for our problem

$$\begin{aligned} L(w, b, \alpha) &= \|w\|^2 + \alpha\|\xi\|^2 + \sum_i {}_i(1 - y_i(w^T x_i + b))^2 \\ &= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i(1 + y_i^2(w^T x_i + b)^2 - 2y_i(w^T x_i + b)) \\ &= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i(1 + y_i^2(\|w\|^2\|x_i\|^2 + b^2 + 2bx_i^T w) - 2y_i(x_i^T w + b)) \\ \frac{\partial L}{\partial w} &= w + \sum_i \alpha_i(y_i^2(2\|x_i\|^2 w + 2bx_i^T) - 2y_i x_i^T) = 0 \\ \sum_i w(\frac{1}{n} + 2\alpha_i(y_i^2\|x_i\|^2) &= \sum_i 2\alpha_i y_i x_i^T(1 - y_i b) \\ w(\frac{1}{n} + 2\alpha_i(y_i^2\|x_i\|^2) &= 2\alpha_i y_i x_i^T(1 - y_i b) \end{aligned} \tag{4}$$

Question3

3a.

Suppose if there is only one point as a support vector, that means all geometric margin are larger then the minimum geometric margin $\gamma_0 = \min \gamma_i = \min y_i(w^T x_i + b)$, except one point (y_0, x_0). Since the hyperplane bisect the data, there is a point on different side such that, for $y_i \neq y_0$, $\gamma_1 = \min \gamma_i = \min y_i(w^T x_i + b)$. Then we have $\gamma_1 + \gamma_0 > 2\gamma_0$, which means the hyperplane is not margin-maximal.

3b.

Yes, there could be more than 2 support vectors. In a 2-D data space, 3 data points which can form a isosceles triangle, then all three points will be the support vectors.