P1:

① $\vec{w} = 0\hat{i} + \hat{j} + \hat{k}$ $(\vec{d} = 0\hat{i} + 2\hat{j} + 2\hat{k})$

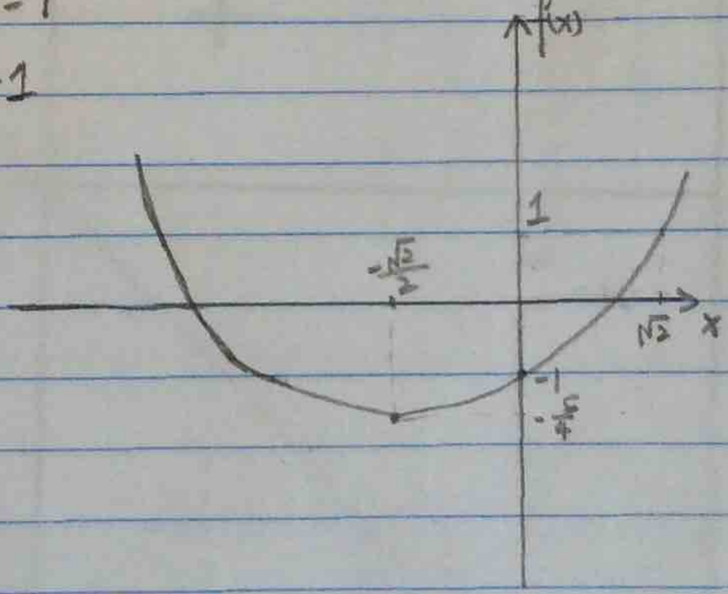
② $\gamma = \frac{1}{2} \cdot \sqrt{0^2+2^2+2^2} = \sqrt{2}$

③ $\gamma = \frac{1}{\|w\|} \Rightarrow w_1^2 + w_2^2 + w_3^2 = (\frac{1}{\gamma})^2 = (\frac{1}{\sqrt{2}})^2$

$w_2 = w_3 = \frac{1}{2}$ $w_1 = 0$. $\hat{w} = [0 \ \frac{1}{2} \ \frac{1}{2}]^T$

④ $\begin{cases} \hat{w}_0 \le -1 - w^T \phi(x_1) = -1 \\ \hat{w}_0 \ge 1 - w^T \phi(x_2) = -1 \end{cases}$

$\therefore \hat{w}_0 = -1$



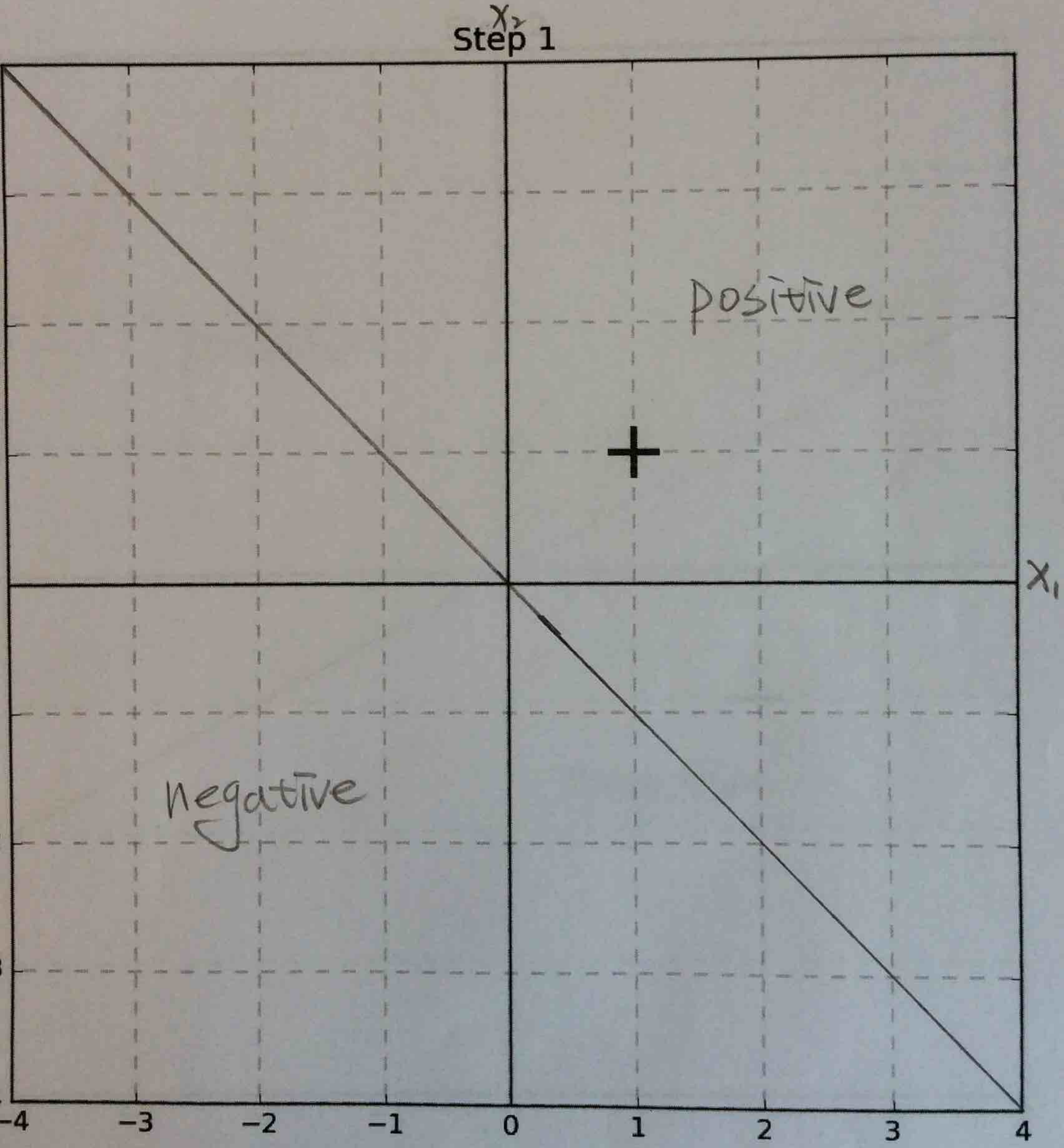
5. $f(x) = -1 + \frac{\sqrt{2}}{2}x + \frac{1}{2}x^2$
$= \frac{1}{2}(x + \frac{\sqrt{2}}{2})^2 - \frac{5}{4}$

P2: ① ② on next page

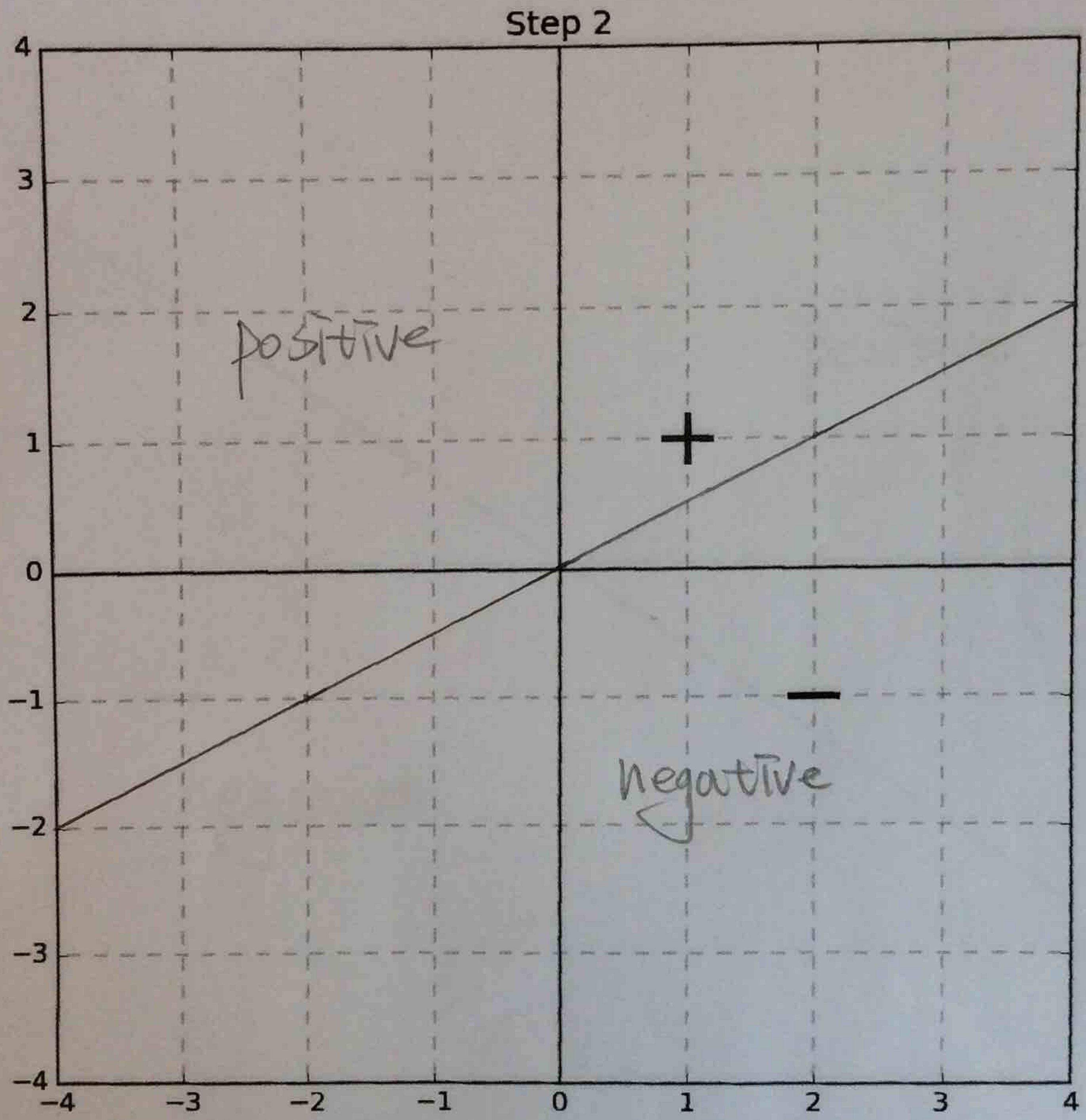
③ mistakes $\le \frac{R^2}{\gamma^2} = \frac{25}{0.5^2} = 100$

$100 - 4 = 96$
96 mistakes allowed for future data

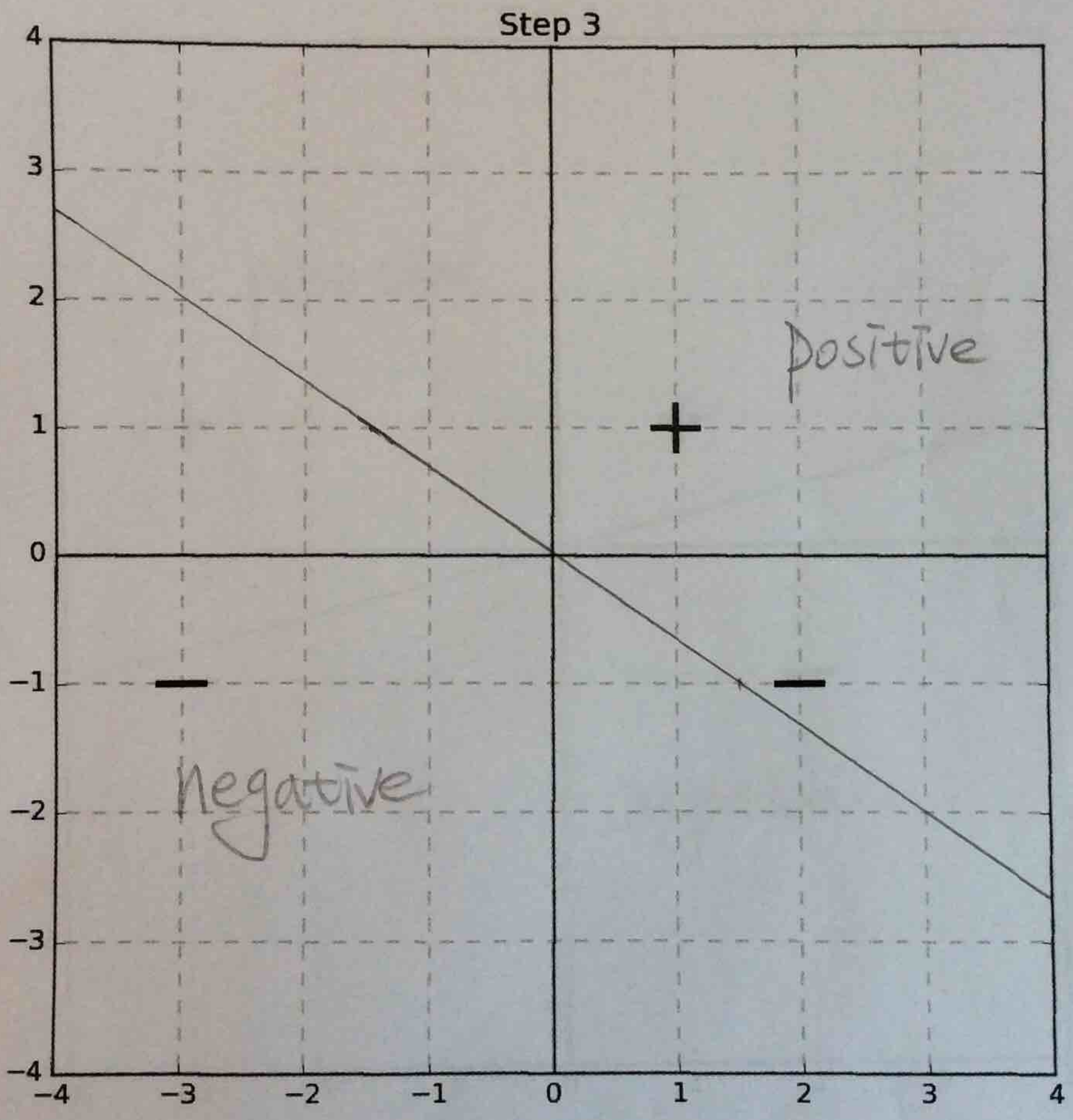
Equation: $X_1 + X_2 = 0$



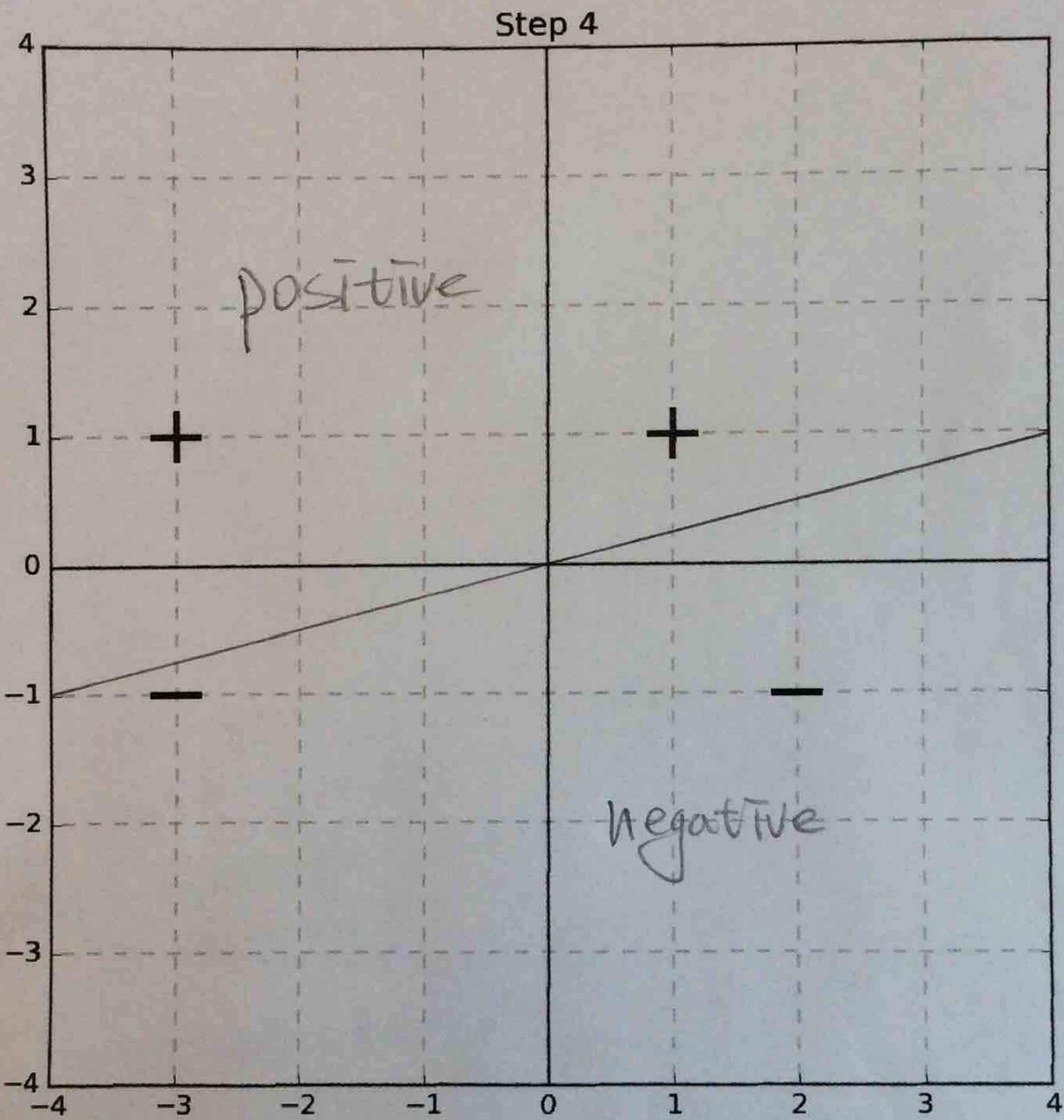
Equation: $x_1 - 2x_2 = 0$



Equation: $2x_1 + 3x_2 = 0$

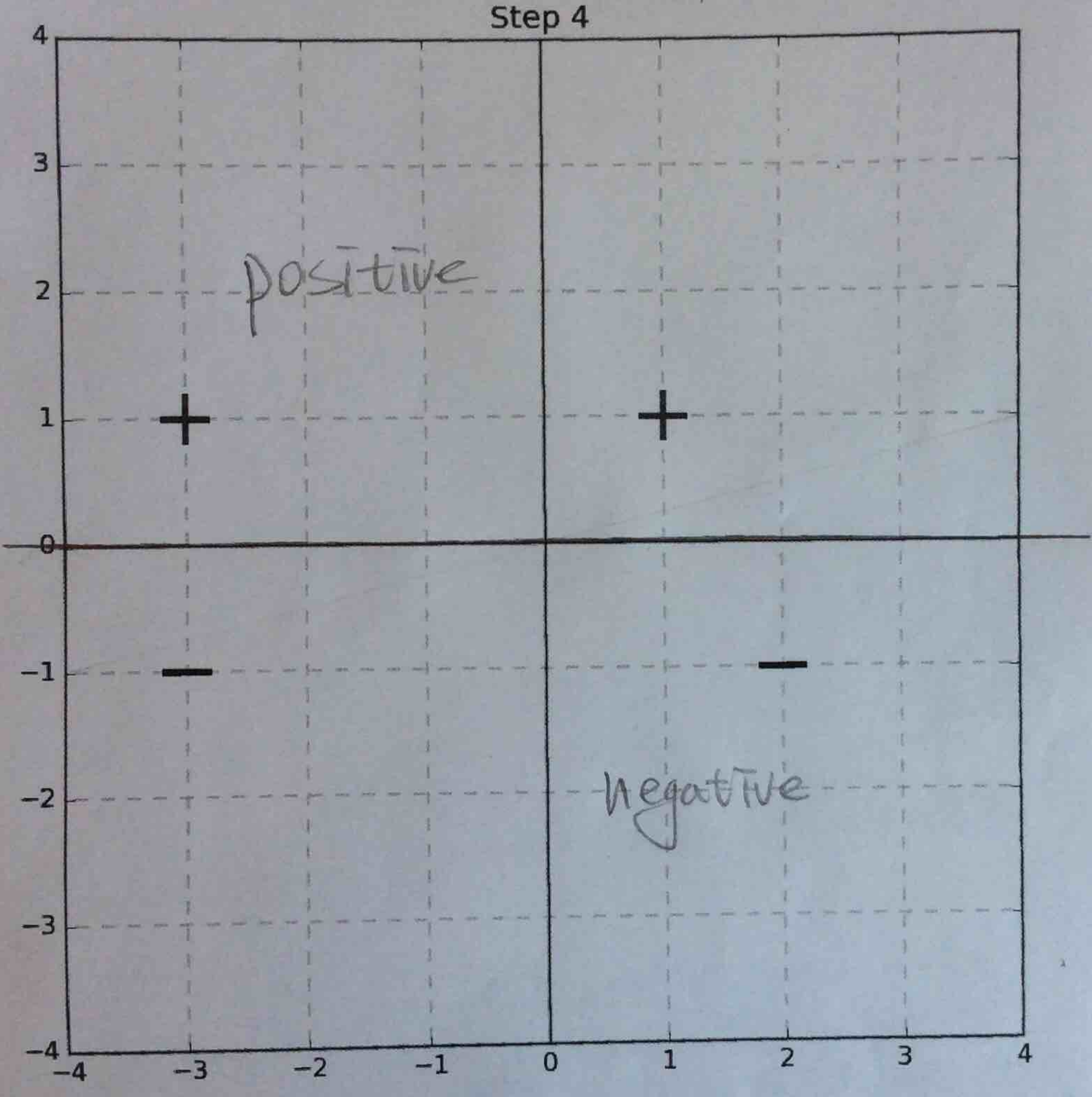


Equation: $x_1 - 4x_2 = 0$



part 1 does not maximize the margin

Equation: $x_2 = 0$



Prat II. Programming

According to Learning rate vs Accuracy graph, we could find that when learning rate =0.1, lammda = 0.001, the accuracy is best. The best accuracy is 0.8393169.