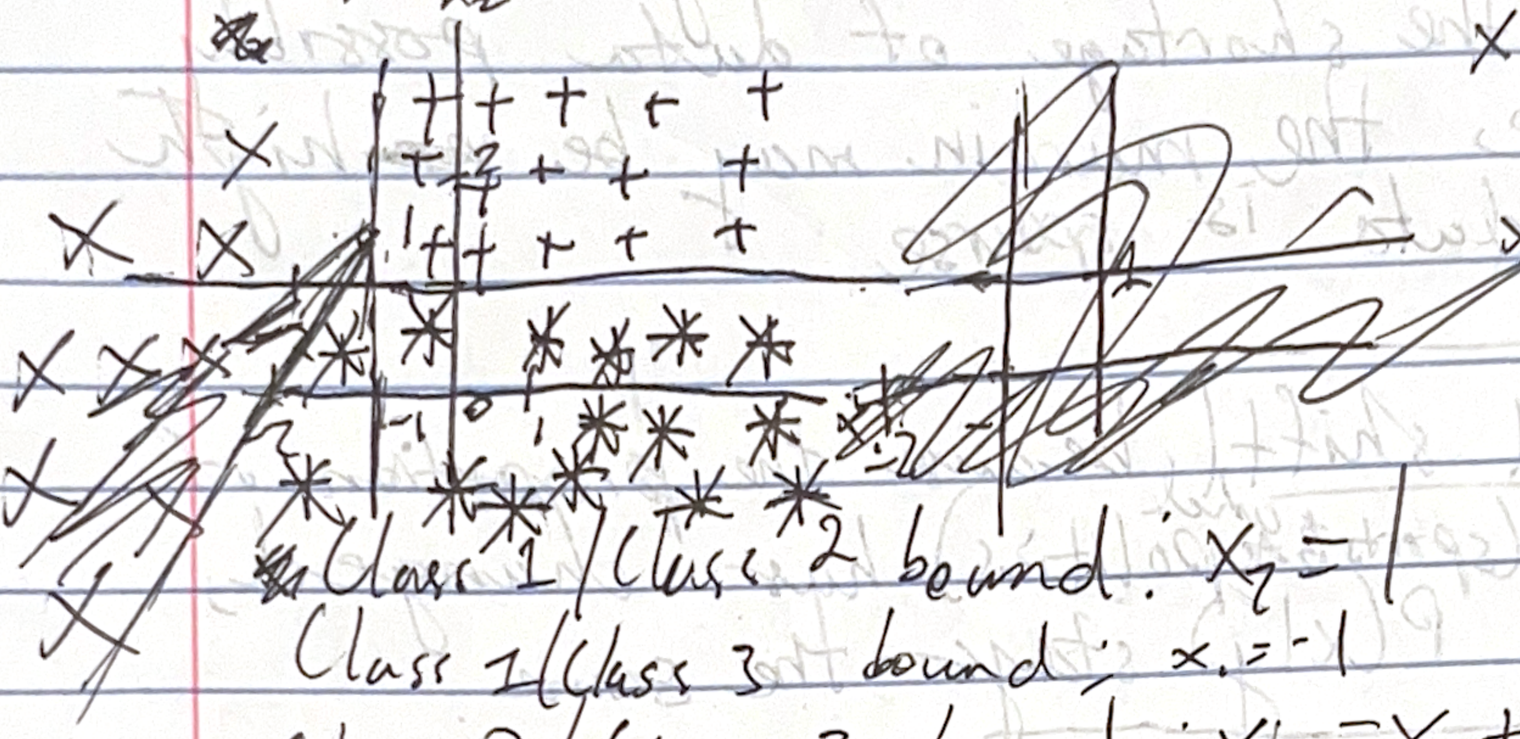
HW #7

1. 1: $x_1 + x_2$ 2: $x_1 + 1$, 3: $x_2 - 1$

Key:
1: +
2: *
3: x



$x_1 + x_2 = x_2 - 1$

$x_1 = -1$

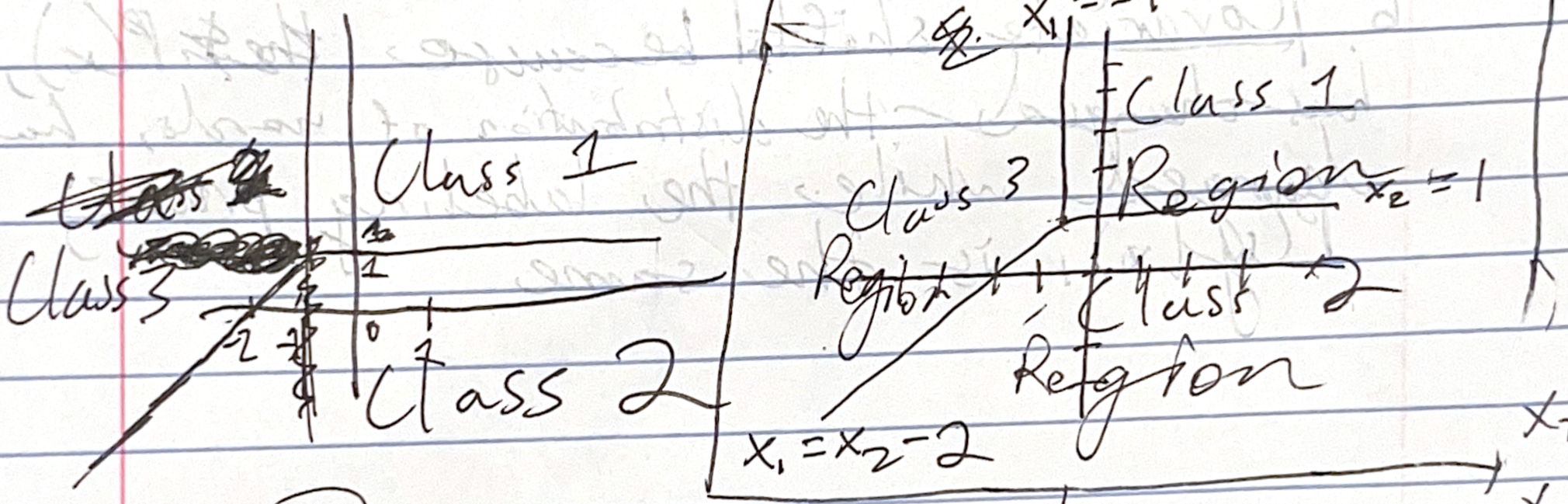
$x_1 + 1 = x_2 - 1$

$x_2 = x_1 + 2$

(1, -2)

$x_2 < 1$ $x_1 > x_2 - 2$

$x_1 = x_2 - 2$



2. a. b. Option b is best because its input most closely matches the task of the model. Option b is best because the photos are the most representative of our input plants from California, so would probably improve the model's performance on new data.

3. We are able to find a good model despite the shortage of data possibly because the margin may be ~~low~~ high or the data is sparse.

4. a. Label shift, because the proportion of labels (sports and politics) has changed, while P(x|y) stays the same.

b. Covariate shift, because P(x), ~~has changed~~, the distribution of words, has changed while the labeling process, P(y|x), is the same.