1. Which of the following is NOT an ensemble method? 1 point

- ○ Gradient boosted trees
- ○ AdaBoost
- ○ Random forests
- ◉ Single decision trees

2. 1 point

Each binary classifier in an ensemble makes predictions on an input x as listed in the table below. Based on the ensemble coefficients also listed in the table, what is the final ensemble model's prediction for x?

	Classifier coefficient w_t	Prediction for x
Classifier 1	0.61	+1
Classifier 2	0.53	-1
Classifier 3	0.88	-1
Classifier 4	0.34	+1

- ○ +1
- ◉ -1

3. (True/False) Boosted trees tend to be more robust to overfitting than decision trees. 1 point

- ○ True
- ◉ False

4. (True/False) AdaBoost focuses on data points it incorrectly predicted by increasing those weights in the data set. 1 point

- ○ True
- ◉ False

5. Let w_t be the coefficient for a weak learner f_t. Which of the following conditions must be true so that $w_t > 0$? 1 point

- ○ $\text{weighted\_error}(f_t) < .25$
- ◉ $\text{weighted\_error}(f_t) < .5$
- ○ $\text{weighted\_error}(f_t) > .75$

- ○ weighted_error(f_t) > .5

6. 1 point

Which of the following classifiers is most accurate as computed on a weighted dataset? A classifier with:

- ◉ weighted error = 0.1
- ○ weighted error = 0.3
- ○ weighted error = 0.5
- ○ weighted error = 0.7
- ○ weighted error = 0.99

7. 1 point

Imagine we are training a decision stump in an iteration of AdaBoost, and we are at a node. Each data point is (x1, x2, y), where x1,x2 are features, and y is the label. Also included are the weights of the data. The data at this node is:

Weight	x1	x2	y
0.3	0	1	+1
0.35	1	0	-1
0.1	0	1	+1
0.25	1	1	+1

Suppose we assign the same class label to all data in this node. (Pick the class label with the greater total weight.) What is the weighted error at the node? Round your answer to 2 decimal places.

0.35

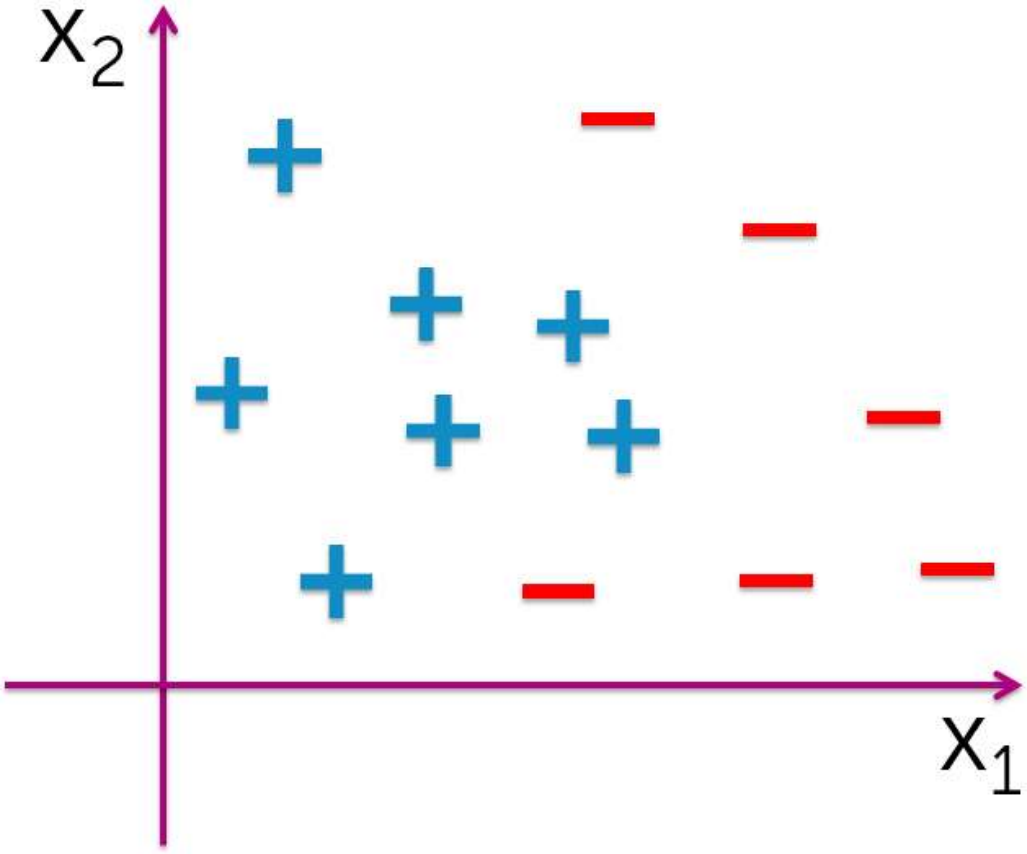
8. 1 point

After each iteration of AdaBoost, the weights on the data points are typically normalized to sum to 1. This is used because

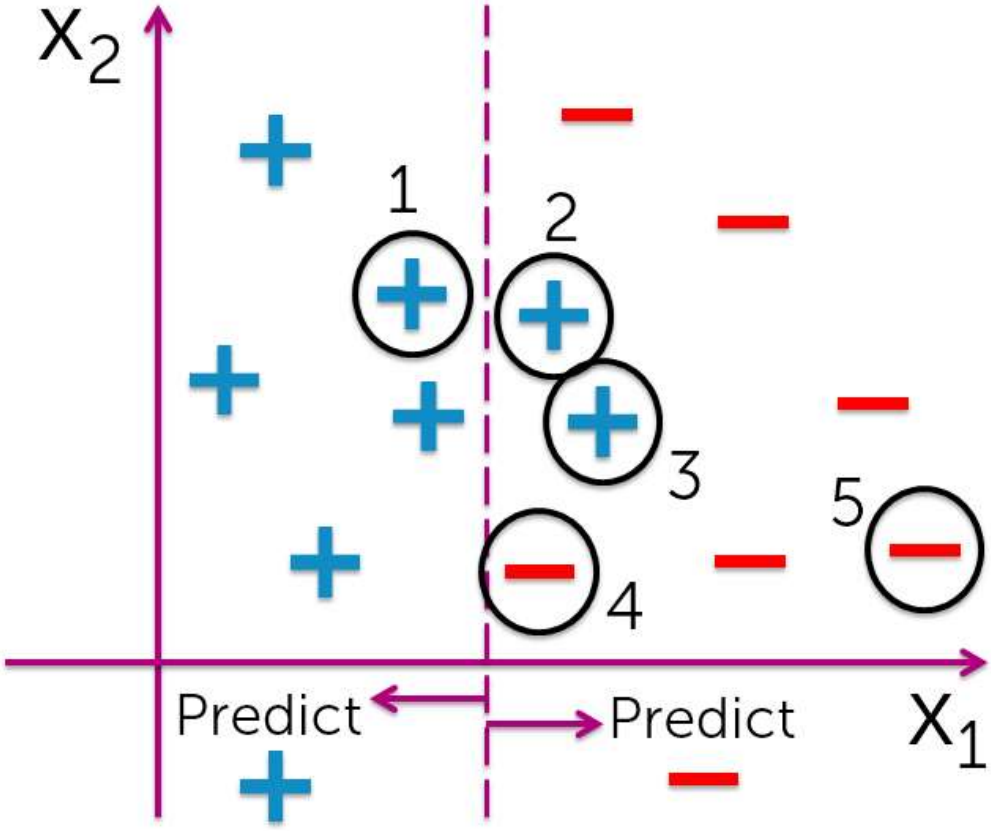
- ◉ of issues with numerical instability (underflow/overflow)
- ○ the weak learners can only learn with normalized weights
- ○ none of the above

9. 1 point

Consider the following 2D dataset with binary labels.



We train a series of weak binary classifiers using AdaBoost. In one iteration, the weak binary classifier produces the decision boundary as follows:



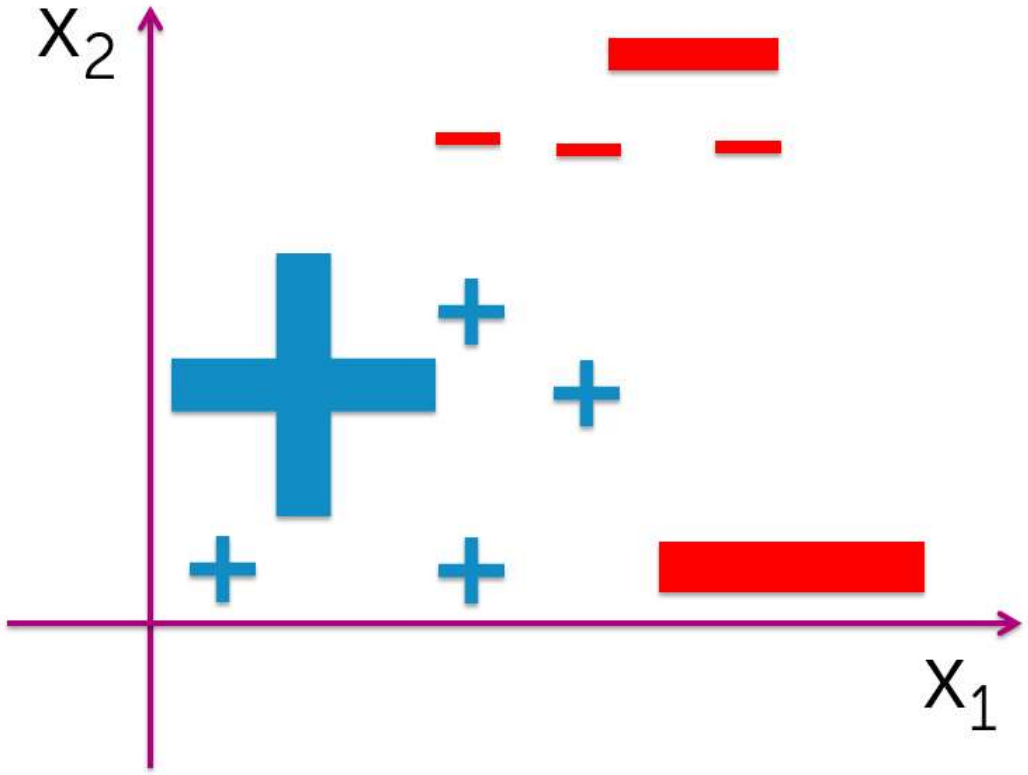
Which of the five points (indicated in the second figure) will receive higher weight in the following iteration? Choose all that apply.

- [ ] (1)
- [x] (2)
- [x] (3)

☐ (4)
☐ (5)

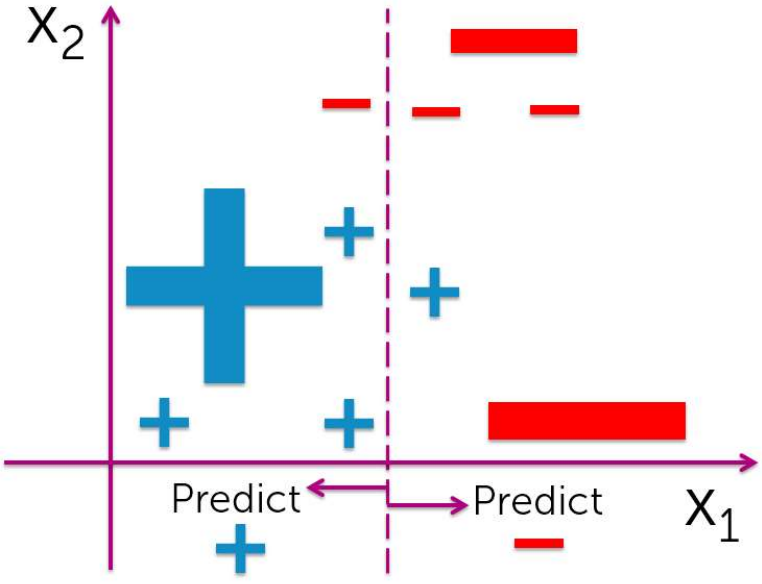
10. 1 point

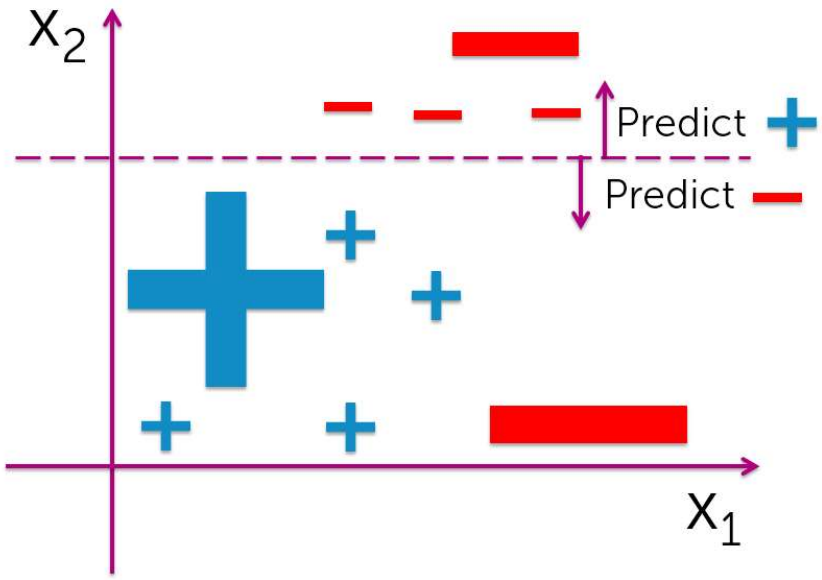
Suppose we are running AdaBoost using decision tree stumps. At a particular iteration, the data points have weights according the figure. (Large points indicate heavy weights.)

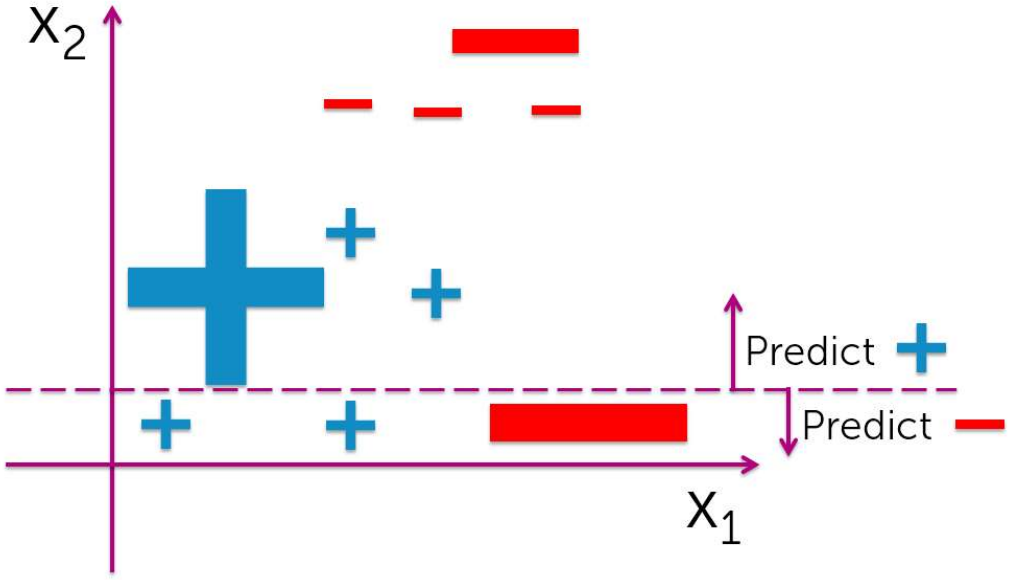


Which of the following decision tree stumps is most likely to be fit in the next iteration?

◉







11. (True/False) AdaBoost achieves zero training error after a sufficient number of iterations, as long as we can find weak learners that perform better than random chance at each iteration of AdaBoost (i.e., on weighted data). 1 point

- (●) True
- ( ) False