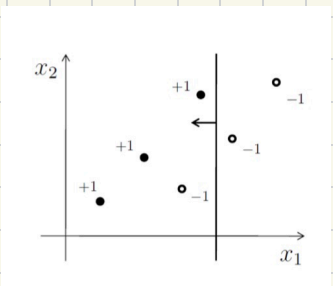
Question 4 - AdaBoost

AdaBoost (ADaptive Boosting) is another approach to the ensemble method field.

It always uses the entire data features (unlike before) and aims to create T weighted classifiers (unlike before, where each classifier had same influence). The new classification will be decided by linear combination of all the classifiers, by:

$$g(x) = \text{sign}\left(\sum_{t=1}^{T} a_t f_t(x)\right), \quad a_t \geq 0$$

Consider the following dataset in $\mathbb{R}^2$:



1) The first decision stump is already drawn, the arrow points in the positive direction. Calculate the classifier error (ε_1) and weight (a_1).

we remember from class that the formula is:

$a_t = \frac{1}{2} \cdot \ln\left(\frac{1 - \text{Error}_t}{\text{Error}_t}\right)$ and because we have only one point that was missclassified and since we know that at the beginning all the points have the same weight then we get:

$\varepsilon_1 = 1/6 \Rightarrow a_1 = \frac{1}{2} \cdot \ln\left(\frac{1-\varepsilon_1}{\varepsilon_1}\right) = \frac{1}{2} \cdot \ln\left(\frac{1-1/6}{1/6}\right) = \frac{1}{2}\ln(5) = 0.804$

2) calculate the new weights of the samples (and normalize them to get valid distribution).

since there was only one missclassified point then all the other points will have the same weight according to the formula: $d_{t+1}(x_i) = d_t(x_i) \cdot \exp(-a_t y_i f_t(x_i))$, we know that $y_i f_t(x_i) \in \{-1, 1\}$ thus for every point that was classified correctly we will have the weight $d = \frac{1}{6} \cdot e^{-0.8} = 0.0745$ and we have 5 points like that. As for the missclassified point we get the weight $d = \frac{1}{6} \cdot e^{0.8} = 0.372$.

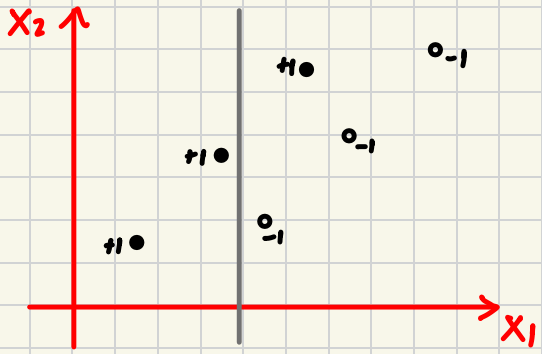
Now to normalize the weights all we have to do is sum the weights and divide them by the result: sum=0.745

missclassified point normalized weight: $0.372/0.745 = 0.499 \approx 0.5$

correctly classified point normalized weight: $0.0745/0.745 = 0.1$

---

3) Draw the second decicion stump. Reminder: the decision stump (our classifiers) are parallel to x/y axis.

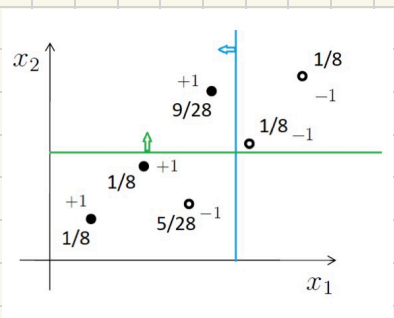


4) without calculations, which classifier's weight is larger a_1 or a_2? Explain why

a_2 is larger meaning that the second classifier has a larger weight because as we can see from the graphs in previous sections, both of the classifiers misclassify the same amount of points, the only difference is the weight for the missclassified point, therefore, since in the second classification we got a missclassified point with a smaller weight than the first classifier then the error will be lower hence we get a larger weight in the second one.

---

5) In the right image, there is the dataset and the weights for each point, after finding the third decision stump and calculating the new weights. which of the following (green or blue) is the correct decision stump?



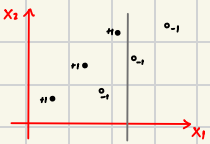
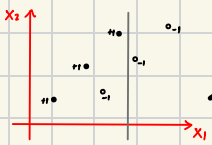
From previous sections we already found the first and second stump, as for the third one, when we update the weights of the points after the second classification, we are supposed to give the missclassified point a larger weight therefore we get the blue classifier and not the green.

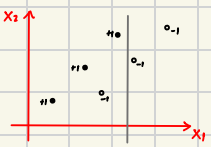
6) Given $a_2 = 1.1$, $a_3 = 0.62$, draw the full classifier, like in slide 13. What is the train accuracy?

similarly to what we did in class, we have:

$a_1 = 0.804$, $a_2 = 1.1$, $a_3 = 0.62$ therefore:

Result = (a_1 × classifier 1 + a_2 × classifier 2 + a_3 × classifier 3)

$\Rightarrow$ (0.804 ×  + 1.1 ×  + 0.62 × )

Result = 

now to calculate the train accuracy we see that only one point was missclassified thus: train-accuracy = $\frac{6-1}{6}\% = 83.33\%$