

**Problem 1.** [20 points] Consider the AdaBoost algorithm we discussed in the class. AdaBoost is an example of ensemble classifiers where the weights in next round are decided based on the training error of the weak classifier learned on the current weighted training set. We wish to run the AdaBoost on the dataset provided in Table 1.

1. Assume we choose the following decision stump  $f_1$  (a shallow tree with a single decision node), as the first predictor (i.e., when training instances are weighted uniformly):

```
if(Color is Yellow):
    predict Edible = Yes
else:
    predict Edible = No
```

What would be the weight of  $f_1$  in final ensemble classifier (i.e.,  $\alpha_1$  in  $f(x) = \sum_{i=1}^K \alpha_i f_i(x)$ )?

2. After computing  $f_1$ , we proceed to next round of AdaBoost. We begin by recomputing data weights depending on the error of  $f_1$  and whether a point was (mis)classified by  $f_1$ . What is the weight of each instance in second boosting iteration, i.e., after the points have been re-weighted? Please note that the weights across the training set are to be uniformly initialized.
3. In AdaBoost, would you stop the iteration if the error rate of the current weak classifier on the weighted training data is 0?

Instance	Color	Size	Shape	Edible	
D1	Yellow	Small	Round	Yes	✓
D2	Yellow	Small	Round	No	✗
D3	Green	Small	Irregular	Yes	✗
D4	Green	Large	Irregular	No	✓
D5	Yellow	Large	Round	Yes	✓
D6	Yellow	Small	Round	Yes	✓
D7	Yellow	Small	Round	Yes	✓
D8	Yellow	Small	Round	Yes	✓
D9	Green	Small	Round	No	✗
D10	Yellow	Large	Round	No	✗
D11	Yellow	Large	Round	Yes	✓
D12	Yellow	Large	Round	No	✗
D13	Yellow	Large	Round	No	✗
D14	Yellow	Large	Round	No	✗
D15	Yellow	Small	Irregular	Yes	✓
D16	Yellow	Large	Irregular	Yes	✓

Table 1: Mushroom data with 16 instances, three categorical features, and binary labels.

$$1) \quad w^0 = \left[ \underbrace{\frac{1}{16}, \dots, \frac{1}{16}}_{16} \right] \quad E_1 = \sum_{i=1}^{16} w_i^0 \mathbb{I}[y_i \neq f_1(x_i)]$$

$$= \left(\frac{1}{16}\right)(0) + \left(\frac{1}{16}\right)(2) + \dots + \left(\frac{1}{16}\right)(0)$$

$$= \left(\frac{1}{16}\right)(6) = \frac{6}{16}$$

$$\alpha_1 = \frac{1}{2} \log\left(\frac{1 - \frac{6}{16}}{\frac{6}{16}}\right) = \frac{1}{2} \log\left(\frac{10/16}{6/16}\right) = \frac{1}{2} \log\left(\frac{5}{3}\right)$$

$$2) \quad w_i^{(k)} = \frac{1}{Z} w_i^{(k-1)} e^{-\alpha_k y_i f_k(x_i)}$$

Right  $\Rightarrow w_i^{(2)} = \frac{1}{Z} w_i^{(1)} e^{-\left(\frac{1}{2} \log\left(\frac{5}{3}\right)\right)(2)(2)}$

$$= \frac{1}{Z} \left(\frac{1}{16}\right) e^{-\frac{1}{2} \log\left(\frac{5}{3}\right)}$$

$$= \frac{1}{16} e^{-0.255} = 0.04843$$

$\frac{1}{Z}$  is just a normalizing constant

Correct so we decrease weight

Wrong  $\Rightarrow w_i^{(2)} = \frac{1}{Z} \left(\frac{1}{16}\right) e^{-\left(\frac{1}{2} \log\left(\frac{5}{3}\right)\right)(-2)(2)}$

$$= \frac{1}{Z} \left(\frac{1}{16}\right) e^{\left(\frac{1}{2} \log\left(\frac{5}{3}\right)\right)}$$

$$= \left(\frac{1}{16}\right) e^{0.255} = 0.08065$$

Incorrect so we increase weight

NEW weight for  
D1, D4, D5, D6, D7, D8, D9, D11, D15, D16  
is 0.04843

NEW weight for  
D2, D3, D10, D12, D13, D14  
is 0.08065

1 Continued...

3) You would stop the iteration since computing the strength  $\alpha_k = \frac{1}{2} \log\left(\frac{1 - \epsilon_k}{\epsilon_k}\right)$  would be undefined. It also implies overfitting.