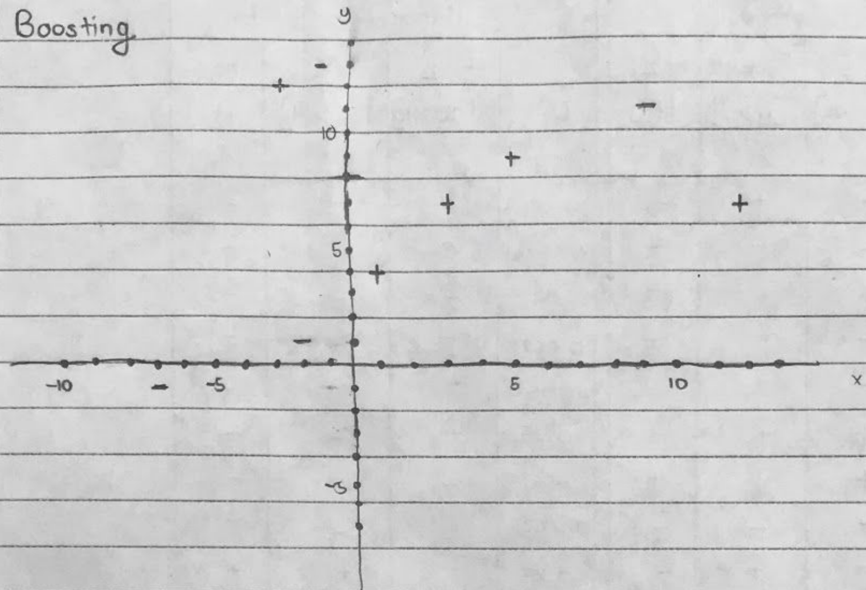
Reinaldo Daniswara
604840665

PSET 4

1 Boosting




a. $D_0 = 0.1$

Best learners $\rightarrow$ $f_1 = [x > 2]$ and $f_2 = [y > 5]$

$\epsilon_{x1} = \frac{2}{10}$ $\quad\quad$ $\epsilon_{x2} = \frac{3}{10}$

$\alpha_0 = \frac{1}{2} \log_2 \frac{1-\epsilon}{\epsilon} = \frac{1}{2} \log_2 \frac{0.8}{0.2} = 1$

c. From the value of $\alpha_0 = 1$ we get

$$D_{t+1}(i) = \frac{D_0(i)}{Z_0} \begin{cases} 2^{-\alpha t} & \text{if } y_i = h_t(x_i) \\ 2^{\alpha t} & \text{if } y_i \neq h_t(x_i) \end{cases}$$

$$D_1(i) = \frac{0.1}{Z_0} \begin{cases} 2^{-1} & \text{if } y_i = h_t(x_i) \\ 2 & \text{if } y_i \neq h_t(x_i) \end{cases}$$

$$= \begin{cases} \frac{1}{20Z_0} & \text{if } y_i = h_t(x_i) \\ \frac{1}{5Z_0} & \text{if } y_i \neq h_t(x_i) \end{cases}$$

$$Z_0 = \frac{8}{20Z_0} + \frac{2}{5Z_0} = 1 \rightarrow \frac{16}{20Z_0} = 1 \rightarrow Z_0 = \frac{4}{5}$$

Thus,

$$D_1(i) = \begin{cases} 0.0625 & \text{if } y_i = h_1(x_i) \\ 0.25 & \text{if } y_i \neq h_1(x_i) \end{cases}$$

Best learners $\rightarrow$ $f_1 = [x > 10]$ and $f_2 = [y > 11]$

continue below the table

Reinaldo Daniswara
604840665

| | | | Hypothesis 1 | | | Hypothesis 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| i | Label | $D_0$ | $f_1 \equiv [x>2]$ | $f_2 \equiv [y>5]$ | $h_1 \equiv [f_1]$ | $D_1$ | $f_1 \equiv [x>10]$ | $f_2 \equiv [y>11]$ | $h_2 \equiv [f_2]$ |
| 1 | − | 0.1 | − | + | − | 0.0625 | − | − | − |
| 2 | − | 0.1 | − | − | − | 0.0625 | − | − | − |
| 3 | + | 0.1 | + | + | + | 0.0625 | − | − | − |
| 4 | − | 0.1 | − | − | − | 0.0625 | − | − | − |
| 5 | − | 0.1 | − | + | − | 0.0625 | − | + | + |
| 6 | − | 0.1 | + | + | + | 0.25 | − | − | − |
| 7 | + | 0.1 | + | + | + | 0.0625 | + | − | − |
| 8 | − | 0.1 | − | − | − | 0.0625 | − | − | − |
| 9 | + | 0.1 | − | + | − | 0.25 | − | + | + |
| 10 | + | 0.1 | + | + | + | 0.0625 | − | − | − |

c. $\epsilon_{f_1} = 1 \times 0.25 + 2 \times 0.0625$

$= 0.3825$

$\epsilon_{f_2} = 0 \times 0.25 + 4 \times 0.0625$

$= 0.25$

$$\alpha_1 = \frac{1}{2} \log_2 \frac{1-\epsilon_{f_2}}{\epsilon_{f_2}} = \frac{1}{2} \log_2 \frac{0.75}{0.25}$$

$$= 0.79$$

d. $H(x) = \text{sgn}\left(1 \times [x>2] + 0.79 \times [y>11]\right)$

Reinaldo Daniswara
604840665

$-\frac{m}{k} - \frac{km+m}{k} =$

2 Multi - class classification

a) i. One vs All $\rightarrow$ k classifiers

All vs All $\rightarrow \binom{k}{2} = \frac{k(k-1)}{2}$

ii. One vs All $\rightarrow$ m examples

All vs All $\rightarrow \frac{2m}{k}$ examples

iii One vs All $\rightarrow$ the rule of winner takes all, or we choose the label that achieves highest score.

All vs All $\rightarrow$ by vote. For ex: apply all vs all classifiers and let each classifier vote the label.

iv. One vs All $\rightarrow O(mk) \rightsquigarrow$ m examples with k classifiers.

All vs All $\rightarrow O\left(\frac{2m}{k} \times \frac{k(k-1)}{2}\right) = O(mk)$.

b). Even though both has the same time complexity, I will choose one vs all because practically it is easier to implement and only has k classifiers. While All vs All need at least $k^2$.

c). Complexity for KERNEL PERCEPTRON $\rightarrow O(m^2)$

Thus, complexity of one vs all using this method, $O(m^2k)$

— " — all vs all $O\left(\frac{4m^2}{k^2} \times \frac{k(k-1)}{2}\right) = O(m^2)$

Therefore, if using this method, all vs all is more preferable than one vs all since $O(m^2) < O(m^2k)$

d) One vs all $\rightarrow O(dm^2k)$

All vs all $\rightarrow O\left(d\,\frac{4m^2}{k^2} \times \frac{k(k-1)}{2}\right) = O(dm^2)$

Here, all vs all is better

e) One vs all $\rightarrow O(d^2mk)$

All vs all $\rightarrow O\left(d^2\,\frac{4m}{k} \times \frac{k(k-1)}{2}\right) = O(d^2mk)$

Since both has the same time complexity, and we don't know the exact

Reinaldo Daniswara
604840665

implementation of this magic box, I would assume they have the same efficiency.

f) Counting $\rightarrow O(m^2)$

since we need to run on each classifier.

$$\frac{m(m-1)}{2} \rightarrow O(m^2)$$

Knockout $\rightarrow O(m)$

Since we don't care about the loser class.