CS146 Winter 2018 - Problem Set 4

Name: Tianyang Zhang
SID: 404743024

1. Boosting

   (a) $D_0 = 0.1$
   Best learners are $f_1 = [x > 2]$ and $f2 = [y > 5]$
   $\epsilon_{x1} = 0.2$ and $\epsilon_{x2} = 0.3$
   $\alpha_0 = \frac{1}{2} log_2 \frac{1-\epsilon}{\epsilon} = \frac{1}{2} log_2 \frac{0.8}{0.2} = 1$

   (b) Answer on the table

   (c) $D_1(i) = \frac{1}{10Z_0} \begin{cases} 2^{-1} & \text{if } y_i = h_t(x) \\ 2^1 & \text{if } y_i \neq h_t(x) \end{cases}$
   The values of $Z_0 = \frac{8}{20Z_0} + \frac{2}{5Z_0} = 1$
   Thus, $D_1(i) = \begin{cases} 0.0625 & \text{if } y_i = h_t(x) \\ 0.25 & \text{if } y_i \neq h_t(x) \end{cases}$
   New best learners are $f_1 = [x > 10]$ and $f2 = [y > 11]$, weighted sum for both cases are $\epsilon_{f1} = 1 \times 0.25 + 4 \times 0.0625 = 0.25$ and $\epsilon_{f2} = 0 \times 0.25 + 2 \times 0.0625 = 0.3825$
   $\alpha_1 = \frac{1}{2} log_2 \frac{1-\epsilon_{f2}}{\epsilon_{f2}} = \frac{1}{2} log_2 \frac{0.75}{0.25} = 0.79$

   (d) $H(x) = sgn(1 \times [x > 2] + 0.79 \times [y > 11])$

		Hypothesis 1				Hypothesis 2			
i (1)	Label (2)	D_0 (3)	$f_1 \equiv [x > 2]$ (4)	$f_2 \equiv [y > 4]$ (5)	$h_1 \equiv [f_1]$ (6)	D_1 (7)	$f_1 \equiv [x > 10]$ (8)	$f_2 \equiv [y > 11]$ (9)	$h_2 \equiv [f_2]$ (10)
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	−	−	−

2. Multi-class classification

   (a) Classifiers

      i. One vs All : k classifiers
      All vs All : $\binom{k}{2} = \frac{k(k-1)}{2}$

      ii. One vs All : m samples
      All vs All : $\frac{2m}{k}$ samples

      iii. One vs All : we choose labels that achieves highest score.
      All vs All: Apply all classifiers and allow those classifiers to vote.

      iv. One vs All : Time complexity O(mk)
      All vs All: O($\frac{2m}{k} \times \frac{k(k-1)}{2}$) = O(mk)

   (b) Since they both has the same time complexity, we need to find other parameters to define which one is better. I think, the implementation of one vs all is easier compare to all vs all, thus I will prefer one vs all method. In addition, one vs all only needs k classifiers, while all vs all needs at least k^2 classifiers.

(c) Since complexity of Kernel Perceptron is $O(m^2)$, it will change our analysis before.
For one vs all it will become $O(m^2k)$
For all vs all it will become $O(\frac{4m^2}{k^2} \times \frac{k(k-1)}{2}) = O(m^2)$
Thus, based on the time complexity that we counted above, we know that all vs all is more poreferable since it is faster that one vs all.

(d) Since complexity of magical black box is $O(dn^2)$, it will change our analysis before.
For one vs all it will become $O(dm^2k)$
For all vs all it will become $O(\frac{4dm^2}{k^2} \times \frac{k(k-1)}{2}) = O(dm^2)$
Thus, based on the time complexity that we counted above, we know that all vs all is more preferable since it is faster that one vs all.

(e) Since complexity of magical black box is $O(d^2n)$, it will change our analysis before.
For one vs all it will become $O(d^2mk)$
For all vs all it will become $O(\frac{4d^2m}{k^2} \times \frac{k(k-1)}{2}) = O(d^2mk)$
Since the time complexity is the same, so their efficiency is the same, and we could not really pick which one is better.

(f) Using **counting** method, in order to do majority vote, we need to run the algorithm on each classifier. Thus,

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

On the other hand, **knockout** method will eliminate loser and only care about the winner, thus, time complexity is $O(m)$