Cody Hubbard 004843389

CSM146, Winter 2018

Problem Set 4

Problem 1

(a) (b) (c) Solution:

i	label	Hypothesis 1				Hypothesis 2			
		D_0	$f_1 \equiv [x > 2]$	$f_2 \equiv [y > 5]$	$h_1 \equiv [x > 2]$	D_1	$f_1 \equiv [x > 6]$	$f_2 \equiv [y > 11]$	$h_1 \equiv [y > 11]$
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	+	+	+	.0625	+	-	-
8	-	0.1	-	-	-	.0625	-	-	-
9	+	0.1	-	+	-	0.25	-	+	+
10	+	0.1	+	+	+	.0625	-	-	-

(c cont.)
Due to the weights on the errors created using h_1 the new best classifier is $f_2 \equiv [y > 11]$

$$\begin{aligned}\alpha_0 &= \frac{1}{2}ln(\frac{1-\epsilon}{\epsilon}) \\ &= \frac{1}{2}ln(\frac{1-\frac{2}{10}}{\frac{2}{10}}) \\ &= \frac{1}{2}ln(\frac{8}{2}) \\ &= ln(4)^{\frac{1}{2}}\end{aligned}$$

so for correctly classified $e^{-\alpha_0} = \frac{1}{e^{ln(4)^{\frac{1}{2}}}} = \frac{1}{2}$ and for incorrectly classified $e^{\alpha_0} = e^{ln(4)^{\frac{1}{2}}} = 2$.

so $D_1 = \frac{D_0}{Z_0} \cdot e^{-\alpha_0}$ and $Z_0 = \frac{1}{10}(8 \cdot \frac{1}{2} + 2 \cdot 2) = 0.8$ then $D_1 = \frac{0.1}{0.8} \cdot 2 = 0.25$ for incorrect points and $D_1 = \frac{0.1}{0.8} \cdot \frac{1}{2} = 0.0625$ for correct points

(d) Solution: $\epsilon_1 = 4 \cdot 0.625 = .25$ and $\alpha_1 = ln(3)^{\frac{1}{2}}$
For Ada-boost $Hfinal(x) = sgn(\sum \alpha_t h_t(x)) = sgn[ln(4)^{\frac{1}{2}} \cdot sgn(x-2) + ln(3)^{\frac{1}{2}} \cdot sgn(y-11)]$

Problem 2

(a) Solution: **(i)** K classifiers for One vs. All and $\dfrac{K(K-1)}{2}$ classifiers of All vs. All

(ii) One vs. All uses m examples and All vs. All uses $\dfrac{2m}{k}$ examples

(iii) I would use argmax for One vs. All and majority vote for All vs. All.

(iv) $O(km)$ for both One vs. All and All vs All

(b) Solution: I prefer All vs. All due to how it is less prone to separability and imbalanced data set problems. Furthermore the method would be learned using a machine so I personally would not see much of the extra computation costs cased by All vs All.

(c) Solution: Using a kernel perception the the complexity for One vs. All becomes $O(km^2)$ and All vs All becomes $O(m^2)$. It doesn't really change my analysis above except that it improves All vs All making my choice even more sensible.

(d) Solution: One vs All becomes $O(kdm^2)$ and All vs. All becomes $O(dm^2)$. All vs All is more efficient.

(e) Solution: One vs All becomes $O(kd^2m)$ and All vs. All becomes $O(d^2m)$. All vs All is more efficient..

(f) Solution: For counting method the complexity is $O(k^2)$ and for knockout it is $O(k)$.