If the best expert makes m mistakes, then

$$W \geq (\frac{1}{2})^m \tag{4.2}$$

Combining inequalities 4.1 and 4.2, it is known

$$n(\frac{3}{4})^M \geq \frac{1}{2}^m$$

$$(4/3)^M \leq n2^m \implies M \log \frac{4}{3} \leq \log n2^m$$

$\implies M \log \frac{4}{3} \leq \log n + m \log 2$
$\implies M \leq \frac{\log n + m \log 2}{\log \frac{4}{3}}$

$$\implies M \leq 2.409441778(m + \log n)$$

For problems where there are three or more outcomes, given a slightly different bound as follows. If there are $c > 2$ outcomes then the outcome must be chosen where the weights of the experts predicting the outcome is at least $\frac{W}{c}$. So, The inequality 4.3 would become

$$W \leq n(\frac{5}{6})^M \tag{4.3}$$

Consequently, the bound would be derived $M \leq \frac{m + \log n}{\log \frac{6}{5}} \leq 3.801784017(m + \log n)$.

Randomized Weighted Majority Algorithm

The “randomized weighted majority algorithm” tightens the error bound of the weighted majority algorithm. In this technique, the final weight of each expert is used to determine the probability for its prediction being chosen as the outcome of the algorithm. Let the elements of the vector of weights $W = w_1, w_2.....$ be initialized to 1 each. Let, once again, $x_1, x_2....$ be the set of opinions given by the experts. Let, for the sake of simplicity $x_i \in \{0, 1\}$. After predictions are made, the correct answer l is revealed. The weights of the mistaken experts are updated by deducting a penalization.

$$\forall\,(x_i \neq l), w_i \leftarrow (1-\beta)w_i$$

The probability of choosing expert opinion x_i is given by

$$P(x_i) = \frac{w_i}{W}$$

Where, $W = \sum_{i=1}^{N} w_i$, denoting the sum of the entire set of weights.

Module four - The Enhancement Algorithm

The mistake bound under the condition that the weight of each of the n experts predicting correctly is rewarded by a factor of α and those predicting wrongly are penalized by a factor of β, is given by

$$M \leq \frac{m\,ln(\frac{1}{\beta}) + (n-m)ln\,(\frac{1}{\alpha}) + ln\,n + (\alpha+2)}{(2-\beta+\alpha)}$$

where, the number of mistakes made by the best expert is given by m.

Let, fraction F_i of experts are mistaken at time instance i. The final weight after time t is given by

$$W \leftarrow W(1-(1-\beta)F_i) + W(1+(1+\alpha))(1-F_i)$$

$$\implies W = n\prod_{i=1}^{t} W(2-(1-\beta)F_i + (1-\alpha)(1-F_i))$$

$$= n\prod_{i=1}^{t}((3+\alpha)-(2-\beta+\alpha)F_i)$$

$$= n\prod_{i=1}^{t}(1-((-\alpha-2)+(2-\beta+\alpha)F_i) \geq \beta^m\alpha^{n-m}$$

Assuming the best expert makes m mistakes, the final weight is in any case greater than the final weight of the best expert. Now, taking log on both sides gives us

$$ln\,n + \sum_{i=1}^{t} ln(1-((-\alpha-2)+(2-\beta+\alpha)F_i) \geq m\,ln(\beta)+$$

$$(n-m)$$

ln(α)Taking negation on both sides gives us,

$$-ln\,n - \sum_{i=1}^{t} ln(1 - ((-\alpha - 2) + (2 - \beta + \alpha)F_i) \leq$$

$$m\,ln(\frac{1}{\beta}) + (n - m)\,ln(\frac{1}{\alpha})$$

$$\implies ln\,n + \sum_{i=1}^{t}(-\alpha - 2) + (2 - \beta + \alpha)F_i \leq m\,ln(\frac{1}{\beta})$$

$$+(n - m)\,ln(\frac{1}{\alpha})$$

Since it is known that, $ln\,(1 - (x + c)) \leq -(x + c)$ as shown in the Figure4.3 below. $ln\,(1 - (x + c))$ is shown in red and $(1 - (x + c))$ in blue line.

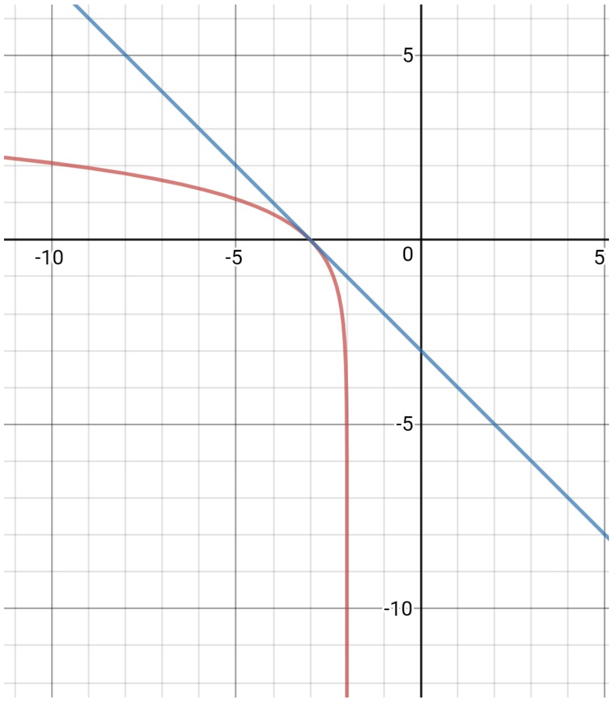

Figure 4.3: Log graph to show $ln\,(1 - (x + c)) \leq -(x + c)$

$$\implies -ln\,n + (-\alpha - 2) + \sum_{i=1}^{t}(2 - \beta + \alpha)F_i \leq m\,ln(\frac{1}{\beta}) +$$

$$(n - m)\,ln(\frac{1}{\alpha})$$

$$or, -ln\,n + (-\alpha - 2) + (2 - \beta + \alpha)M \leq m\,ln(\frac{1}{\beta}) +$$

$$(n - m)\,ln(\frac{1}{\alpha})$$

Where M is the total number of mistakes made by the algorithm, and $M = \sum_{i=1}^{t} F_i$. Rearranging, it is known

$$M \leq \frac{m\,ln(\frac{1}{\beta}) + (n-m)ln\,(\frac{1}{\alpha}) + ln\,n + (\alpha+2)}{(2-\beta+\alpha)}$$

It is to be noted that this bound value is significantly lesser than the conventional method, since in the denominator a factor of $(2+\alpha)$ is received which reduces the overall bound value by at least 50%, since β value in usual cases does not exceed 0.5. However, the exact percentage of tightening is simply an assumption. Please refer to Appendix II for the programming code and results.

4.2.4 Performance Analysis

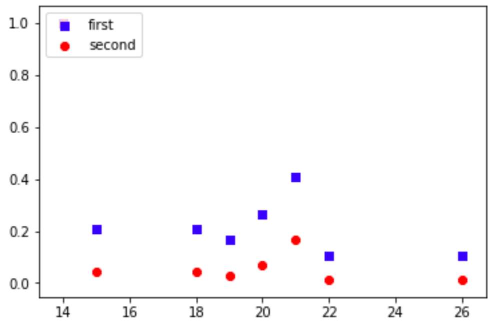

Figure 4.4: The above figure shows how the weights change across the training and test sets. The blue data-points represent the weight reduction in the first training set, as compared to that in the second training set (data-points marked in red)

The set of weights W after first training set is given as:

[0.10737418240000003, 0.10737418240000003, 0.4096, 0.26214400000000004, 0.16777216000000003, 0.20971520000000005, 0.20971520000000005, 1.0]

.

Similarly, set of W weights after second training is given as:

[0.01152921504606847, 0.01152921504606847, 0.16777216000000003, 0.06871947673600001, 0.028147497671065603, 0.043980465111040010, 0.04398046511104001, 1.0]

As can be seen from the above two graphs, when compared, the optimized method 4.7 yields a more comprehensive way of examining the best expert opinion, rather than the randomized approach, 4.6 which yields, in the later training stages, a less comprehensible and chaotic