Notes 6: Weighted Majority

1. Online regret bound model

- e.g. stock market prediction: guessing whether it will go up or down for each day
- A sequence of rounds/trials, each being:
  1. A new unlabeled example $x$ arrives
  2. $n$ experts reveal their opinions about the label for $x$ (label is either 0 or 1)
  3. Algorithm predicts 0 or 1 according to experts’ opinions
  4. Algorithm is told correct label for $x$

- Goal: minimize number of mistakes, compared with the best expert

If every “expert” makes many mistakes, algorithm may, too

2. Weighted Majority

- Fix parameter $0 \leq \beta < 1$
- Initialize: $w_1 = \cdots = w_n = 1$
- On input $x$, poll opinions from experts
  - Compute total weight $q_0$ of experts predicting 0 and total weight $q_1$ predicting 1
  - Predict according to weighted majority (predict 0 if $q_0 > q_1$; predict 1 otherwise)
- On revealing correct label, penalize incorrect experts
  - Multiply every incorrect expert $i$’s weight $w_i$ by $\beta$

If $\beta = 0$, Weighted Majority algorithm becomes Halving algorithm

expert $\leftrightarrow$ concept

expert $i$’s opinion in $j$th trial $\leftrightarrow$ concept $c$’s classification for $j$th sample

No longer assume any expert/concept correctly classifies all samples

Robust to classification noise

**Theorem 2.1.** For any trial sequence, if the best expert (out of $n$ experts) makes $m$ mistakes, then number of mistakes of Weighted Majority is at most

$$\frac{\log n + m \log(1/\beta)}{\log(\frac{2}{1+\beta})}$$

- e.g. $\beta = 1/2$: $2.41(m + \log n)$
- e.g. $\beta = 3/4$: $2.2m + 5.2 \log n$
- e.g. $\beta = 1 - \varepsilon$: $\approx (2 + \frac{3}{2}\varepsilon)m + \frac{2}{\varepsilon} \log n$

**Proof.** Let $W = q_0 + q_1$ = total weight of all experts (initially $n$)

After each mistake, at least half of $W$ shrinks by factor $\beta$

Total weight reduces to $\leq \frac{W}{2} + \frac{W\beta}{2} = \frac{1+\beta}{2}W$

when Weighted Majority makes $M$ mistakes: $W \leq (\frac{1+\beta}{2})^M n$

when best expert makes $m$ mistakes: $w_i = \beta^m$

$w_i \leq W \quad \Rightarrow \quad \beta^m \leq (\frac{1+\beta}{2})^M n \quad \iff \quad m \log \beta \leq M \log(\frac{1+\beta}{2}) + \log n$

$\iff \quad M \log(\frac{2}{1+\beta}) \leq \log n + m \log(1/\beta)$

**Note:** The bound can be interpreted as

$$\frac{\log(W_{\text{init}}/W_{\text{final}})}{\log(1/u)}$$

where $u = \frac{1 + \beta}{2}$ = shrink in $W$ per mistake
Randomized Weighted Majority

Fix parameter $0 \leq \beta < 1$

Initialize: $w_1 = \cdots = w_n = 1$

On input $x$, poll opinions from experts

Predict according to a random expert $i$ chosen with probability proportional to $w_i$

i.e. probability $w_i/W$, where $W =$ total weight $= \sum_{1 \leq i \leq n} w_i$

On revealing correct label, penalize incorrect experts

Multiply every incorrect expert $i$’s weight $w_i$ by $\beta$

Denote $\varepsilon = 1 - \beta$

**Theorem 3.1.** Given any trial sequence with fixed correct labels, if the best expert (out of $n$ experts) makes $m$ mistakes, then

$$\mathbb{E}[\# \text{mistakes of RWM}] \leq \frac{\ln n - m \ln (1 - \varepsilon)}{\varepsilon}$$

e.g. $\beta = 1/2$: $1.39m + 2 \ln n$

e.g. $\beta = 3/4$: $1.16m + 4 \ln n$

e.g. $\beta = 1 - \varepsilon$: $\approx (1 + \frac{\varepsilon}{2})m + \frac{1}{\varepsilon} \ln n$

Key benefit: $\approx m$ mistakes (ignoring additive $\log n$), down from $\approx 2m$

**Proof.** Fix any sequence of $T$ trials together with their correct labels

Let $F_t =$ fraction of total weight on wrong prediction at trial $t$

Want to bound $\mathbb{E}[\# \text{mistakes of RWM}] = \sum_{1 \leq t \leq T} F_t$

At trial $t$, probability of mistake is $F_t$, and $\varepsilon F_t$ fraction of weight is removed

$$W_{\text{final}} = W_{\text{init}}(1 - \varepsilon F_1) \cdots (1 - \varepsilon F_T) \quad (W_{\text{init}} = n)$$

$$\ln W_{\text{final}} = \ln n + \ln (1 - \varepsilon F_1) + \cdots + \ln (1 - \varepsilon F_T)$$

Best expert makes $m$ mistakes:

$w_i = \beta^m = (1 - \varepsilon)^m$

$$W_{\text{final}} \geq w_i \iff \ln W_{\text{final}} \geq \ln w_i \iff \ln n + \sum_{1 \leq t \leq T} \ln (1 - \varepsilon F_t) \geq m \ln (1 - \varepsilon)$$

Claim: $\ln (1 - x) \leq -x$ for all $x$

Take $x = \varepsilon F_t$ in Claim, we get $\ln (1 - \varepsilon F_t) \leq -\varepsilon F_t$, and

$$\varepsilon \sum_{1 \leq t \leq T} F_t \leq \sum_{1 \leq t \leq T} -\ln (1 - \varepsilon F_t) \leq \ln n - m \ln (1 - \varepsilon) \qed$$

Above Claim is true because for all real $x$

$$1 - x \leq e^{-x}$$