4 Nov 2024 Introducing Multiplicative Weights

The HALVING algorithm for sequential prediction.

GIVEN: "example space", a set X.

"hypothesis class", a set $\mathcal{H} = \{h : X \to \{0,1\}\}$,

$\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$ finite.

In each of T "prediction rounds":

1. Example $x_t \in X$ presented to learner.
2. Learner must predict $p_t \in \{0,1\}$.
3. Label $y_t \in \{0,1\}$ is revealed.

Penalty is $\mathbb{1}[p_t \neq y_t] = \begin{cases} 1 & \text{if } p_t \neq y_t \\ \emptyset & \text{if } p_t = y_t \end{cases}$.

Theorem: There is a (deterministic) algorithm for computing p_t given knowledge of $\{h_1, \ldots, h_n\}$ and $\{(x_i, y_i)\}_{i=1}^{t-1}$ such that if $\exists h \in \mathcal{H}$ with $h(x_t) = y_t$ $\forall t$ then $\sum_{t=1}^{T} \mathbb{1}[p_t \neq y_t] \leq \lfloor \log_2(n) \rfloor$.

ALGORITHM. Define $\mathcal{H}_t = \{h \in \mathcal{H} \mid h(x_s) = y_s \ \forall s < t\}$.

Each element in $\mathcal{H}_t$ votes on whether $y_t = 0$ or $y_t = 1$.

Predict p_t according to majority vote.

Observation: Each time $p_t \neq y_t$,

$$|\mathcal{H}_{t+1}| \leq \frac{1}{2} |\mathcal{H}_t|.$$

If we assume $\exists\, h \in \mathcal{H}$ that satisfies $h(x_t) = y_t \;\forall t$, $|\mathcal{H}_T| \geq 1$ because $h \in \mathcal{H}_T$.

Since every mistake decreases $\log_2 |\mathcal{H}_t|$ by at least 1,

$$\#\text{ mistakes} \leq \lfloor \log_2(n) \rfloor.$$

The WEIGHTED MAJORITY algorithm.

Each $h_i \in \mathcal{H}$ has a weight w_i.

At time t, if h_i made $m_i(t)$ mistakes on $x_1, \ldots, x_{t-1}$, its weight is

$$w_i(t) = (1-\varepsilon)^{m_i(t)}$$

$$e^{-\varepsilon} = 1 - \varepsilon + \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{6} + \ldots$$

$e^{-\varepsilon} \geq 1 - \varepsilon$ always.

$$(1-\varepsilon)^x \leq 1 - \varepsilon x \quad 0 \leq x \leq 1$$

Predict p_t according to weighted-majority vote of $\mathcal{H}$.

<u>Analysis</u>: If $p_t \neq y_t$ it means

$$\sum_{i: h_i(x_t) = y_t} w_i(t) \leq \sum_{i: h_i(x_t) \neq y_t} w_i(t).$$

If $W_t = \sum_{i=1}^{n} w_i(t)$

wherever WMAJ makes a mistake

$$W_{t+1} \leq \left(1 - \frac{\varepsilon}{2}\right) \cdot W_t.$$

Suppose WMAJ makes m mistakes in total

and some $h_i \in \mathcal{H}$ makes only m^* mistakes.

$$(1-\varepsilon)^{m^*} < W_{T+1} \leq \left(1 - \frac{\varepsilon}{2}\right)^m W_0 = \left(1 - \frac{\varepsilon}{2}\right)^m \cdot n$$

$w_i(T+1)$

$$(1-\varepsilon)^{m^*} \leq \left(1 - \frac{\varepsilon}{2}\right)^m \cdot n$$

$$m^* \ln(1-\varepsilon) < m \ln\left(1 - \frac{\varepsilon}{2}\right) + \ln(n)$$

$$m < \frac{-\ln(1-\varepsilon)}{-\ln(1-\varepsilon/2)} \cdot m^* + \frac{\ln(n)}{-\ln(1-\varepsilon/2)}$$

$(0 < \varepsilon < 1)$ $\quad < 2(1+\varepsilon) m^* + \frac{2}{\varepsilon} \ln(n).$