Two algorithms:

1) Multiplicative Weight Update (a.k.a Experts Algorithm, a.k.a Hedge Algorithm)

2) Johnson-Lindenstrauss (for dimensionality reduction)

Multiplicative Weight update

Setup

- Imagine you have n experts that each day give you advice on a decision you have to make.

- Each day you listen to an expert (your decision could be probabilistic, i.e. a distribution over experts)

- After that, you find out which expert was right or wrong.

Questions:
- How often you follow the wrong advice?
- Can you hope to do as well as the best expert?

Formal Setup

- There are n possible strategies or experts

- At each time $t = 1, 2, \dots$ we are required to have an array: $(x_1^{(t)}, \dots, x_n^{(t)})$ where $x_i^{(t)} \in [0,1]$ and $\sum_{i=1}^{n} x_i^{(t)} = 1$; $x_i^{(t)}$ is the probability we follow expert i advice.

- After we make our choice, a loss array

$$\ell^{(\tau)} = (\ell_1^{(\tau)}, \dots, \ell_n^{(\tau)}),$$

where $\ell_i^{(\tau)}$ specifies the loss of following expert i at time τ.

- We can think of $\ell_i^{(\tau)} \in [0,1]$ with $\ell_i^{(\tau)} = 0$ meaning no loss.

- Expected loss at time τ:

$$L_\tau = \sum_{i=1}^{n} x_i^{(\tau)} \cdot \ell_i^{(\tau)}$$

- Total Expected Loss in T days:

$$L(T) = \sum_{\tau=1}^{T} L_\tau = \sum_{\tau=1}^{T} \sum_{i=1}^{n} x_i^{(\tau)} \ell_i^{(\tau)}$$

- Consider the loss incurred by the best expert:

$$\min_{i=1,\dots,n} \sum_{\tau=1}^{T} \ell_i^{(\tau)}$$

- We aim for an algorithm that minimizes:

$$R(T) = L(T) - \min_{i=1,\dots,n} \sum_{\tau=1}^{T} \ell_i^{(\tau)}$$

↳ This is called "the regret"

MWU-Algorithm

- Let $\varepsilon \in (0, 1/2)$ (to be choosen later)
- Let $w^{(1)} = (1, \ldots, 1)$

for $\tau = 1$ to T

- Choose expert i with probability:
$$x_i^{(\tau)} = \frac{w_i^{(\tau)}}{\sum_{i=1}^{n} w_i^{(\tau)}}$$
- Set $w_i^{(\tau+1)} = w_i^{(\tau)} (1-\varepsilon)^{\ell_i(\tau)}$ for each $i = 1, \ldots, n$

Idea: Big loss ⇒ that the corresponding weight is decreased by a large amount.

Theorem: $$R(T) \leq \varepsilon T + \frac{\ln n}{\varepsilon}$$

In particular, if $T > 4 \ln n$ and $\varepsilon = \sqrt{\frac{\ln n}{T}}$

$$R(T) \leq 2\sqrt{T \ln n}.$$

$\Rightarrow$ So if we look at "regret-per-day" this is:

$$\frac{R(T)}{T} \leq 2\sqrt{\frac{\ln n}{T}} \longrightarrow 0 \quad \text{as } T \to \infty.$$

$\hookrightarrow$ Very strong guarantee!!

Proof of Correctness

Let $W_t = \sum_{i=1}^{n} w_i^{(t)}$, $L^* = \min_{i=1,\dots,n} \sum_{t=1}^{T} \ell_i^{(t)}$,

$$i^* = \arg\min_{i=1,\dots,n} \sum_{t=1}^{T} \ell_i^{(t)}$$

Let J be any expert. Then:

$$W_{T+1} = \sum_{i=1}^{n} w_i^{(T+1)} \geq w_J^{(T+1)} = \prod_{t=1}^{T} (1-\varepsilon)^{\ell_J^{(t)}} = (1-\varepsilon)^{\sum_{t=1}^{T} \ell_J^{(t)}}$$

$$\boxed{W_{T+1} \geq (1-\varepsilon)^{L^*}}$$

(If W_{T+1} small, then offline optimum L^* must be large)

We claim next that: $W_{t+1} \leq W_t (1 - \varepsilon L_t)$

$\hookrightarrow$ expected loss at time t.

$$W_{t+1} = \sum_{i=1}^{n} w_i^{(t+1)} = \sum_{i=1}^{n} (1-\varepsilon)^{\ell_i^{(t)}} \cdot w_i^{(t)} \leq \sum_{i=1}^{n} \left(1 - \varepsilon \cdot \ell_i^{(t)}\right) \cdot w_i^{(t)}$$

The last inequality follows from: $(1-\varepsilon)^x \leq 1 - \varepsilon x$ for $\varepsilon, x \in [0,1]$

$$W_{\tau+1} \le W_\tau \sum_{i=1}^{n} x_i^{(\tau)} \left(1 - \varepsilon \ell_i^{(\tau)}\right) = W_\tau \left(1 - \varepsilon L_\tau\right) \quad \text{(as claimed).}$$

So, $W_{T+1} \le W_T (1 - \varepsilon L_T) \le W_{T-1} (1 - \varepsilon L_T)(1 - \varepsilon L_{T-1}) \ldots$

$$W_{T+1} \le n \prod_{\tau=1}^{T} (1 - \varepsilon L_\tau)$$

So,

$$(1-\varepsilon)^{L^*} \le W_{T+1} \le n \cdot \prod_{\tau=1}^{T} (1 - \varepsilon L_\tau)$$

$$L^* \cdot \ln(1-\varepsilon) \le \ln n + \sum_{\tau=1}^{T} \ln(1 - \varepsilon L_\tau)$$

Since: $-z - z^2 \le \ln(1-z) \le -z$

$$L^*(-\varepsilon - \varepsilon^2) \le \ln n - \varepsilon \sum_{\tau=1}^{T} L_\tau$$

$$\sum_{\tau=1}^{T} L_\tau - L^* \le \frac{\ln n}{\varepsilon} + \varepsilon L^*$$

$$R(T) \le \varepsilon T + \frac{\ln n}{\varepsilon}$$, as claimed

■

Dimensionality Reduction

Th 1 [Johnson-Lindenstrauss '84] (JL)

For any set X of n points or vectors in $\mathbb{R}^d$, and for all $\varepsilon \in (0,1)$, $\exists$ a mapping $\varphi : \mathbb{R}^d \to \mathbb{R}^k$ where $k = \frac{24 \log n}{\varepsilon^2}$

s.t. $\forall x, y$ in X:

$$(1-\varepsilon)\|x-y\|_2^2 \leq \|\varphi(x)-\varphi(y)\|_2^2 \leq (1+\varepsilon)\|x-y\|_2^2$$

$$\|u-v\|_2^2 = \sum_{i=1}^{d} (u_i - v_i)^2.$$ (ℓ_2, 2-norm, etc.)

Application 1 All pairs distances:

$$O(n^2 d) \quad \text{vs} \quad O\left(n^2 \cdot \frac{\log n}{\varepsilon^2}\right).$$

$\Rightarrow$ There are many ways of choosing φ.

We define φ using standard normal (i.e., Gaussian) random variables.

Let

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & & \vdots \\ A_{d1} & \cdots & A_{dk} \end{pmatrix}_{d \times k} \qquad A_{ij} \sim N(0, 1/k).$$

Define $\varphi(x) = A \cdot x$.

If $x \in \mathbb{R}^d$, then $Ax \in \mathbb{R}^k$.

Normal r.v. review

$$X \sim N(\mu, \sigma^2)$$

μ: mean

σ: std deviation.

PDF: $$n(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}.$$

i.e $\Longrightarrow$ $$\Pr[a \le X \le b] = \int_a^b n(x)\,dx.$$

Properties

(1) $X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$X_1 + X_2 \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.

(2) $\alpha \cdot X_1 = N(\alpha \cdot \mu_1,\ (\alpha \sigma_1)^2)$.

(3) $\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i,\ \sum_{i=1}^{n} (\alpha_i \sigma_i)^2\right)$.

$X_i \sim N(\mu_i, \sigma_i^2)$

(4) Concentration $\quad X_1, \ldots, X_n \sim N(0,1)$ i.i.d's.

$$P_1\left[\left|\sum_{i=1}^{n} X_i^2 - n\right| > \delta n\right] \le 2\exp\left[-\frac{\delta^2 n}{8}\right].$$

for any $\delta \in (0,1)$.

(*) We will prove this concentration bound later.

Proof of JL:

We will show that $\forall x$:

$$P_r\left[(1-\varepsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\varepsilon)\|x\|_2^2\right] \geq 1 - \frac{2}{n^3}.$$

$\circledast$ Think of $x = u - v$ $\quad Ax = Au - Av = \varphi(u) - \varphi(v).$

$\Rightarrow$ The result then follows from a union bound over all pairs in X (there are $\binom{n}{2} \sim \frac{n^2}{2}$ of them).

$\Rightarrow \underbrace{A}_{k \times d}x$ is a k-dimensional vector. $\quad (\ )_{k \times d}$

$$\rightarrow (Ax)_i = \sum_{j=1}^{d} A_{ij} \cdot x_j = Y_i \sim N\left(0, \frac{\sum_j x_j^2}{k}\right).$$

$$= N\left(0, \frac{\|x\|_2^2}{k}\right).$$

$$\|Ax\|_2^2 = \sum_{i=1}^{k} (A \cdot x)_i^2 = \sum_{i=1}^{k} Y_i^2$$

$\Rightarrow$ We renormalize to make the r.v.'s $N(0,1)$

$$Z_i^2 = Y_i^2 \cdot \frac{k}{\|x\|_2^2} \Rightarrow Z_i = \frac{\sqrt{k}}{\|x\|_2^2} \cdot Y_i \Rightarrow Z_i \sim N(0,1)$$

$$\Pr\left[\|Ax\|_2^2 \geq (1+\varepsilon)\|x\|_2^2\right] = \Pr\left[\sum_{i=1}^{k} y_i^2 \geq (1+\varepsilon)\cdot\|x\|_2^2\right]$$

$$= \Pr\left[\sum_{i=1}^{k} z_i^2 \geq (1+\varepsilon)k\right] \leq 2\exp\left(-\frac{\varepsilon^2 \cdot k}{8}\right)$$

$$k = \frac{24 \ln n}{\varepsilon^2}$$

$$\leq \frac{2}{n^3}$$ as claimed $\blacksquare$