Week 5 — Lecture 25

$\forall\, u \in \mathbb{R}^d$

1) $R(u) = \frac{1}{2\eta} \|w\|_2^2$

2) $R(u) = \sum_{i=1}^{d} u_i \log u_i$

$u \in S = \{ w,\ w_i \geq 0,\ \underbrace{\sum w_i}_{\|w\|_1} \leq B \}$

$\langle \nabla^2 R(w)\, x, x \rangle \geq \sigma \|x\|^2$

$\frac{1}{B}$ — strong convex w.r.t. l_1-norm

Regret with Euclidean Regularizer

Assume: $f_1, f_2, \ldots f_n$ are L-Lipschitz w.r.t l_2-norm

Applying FoReL :

$$\text{Regret}(n, u) \leq R(u) - \cancel{\min_u R(u)}^{\,0} + \frac{nL^2}{\sigma} \qquad \boxed{\sigma = \frac{1}{\eta}}$$

$$\leq \frac{1}{2\eta}\|u\|_2^2 + n\eta L^2$$ L^2

Let $\|u\|_2 \leq B$

$$\therefore \quad \leq \frac{1}{2\eta} B^2 + n\eta L^2$$

$$\leq BL\sqrt{2n} \quad \text{by optimizing over } \eta$$

With Euclidean regularizer the update rule:

$w_{t+1} = w_t - \eta z_t$ holds

Thus redefine $R(u)$

$$R(u) = \begin{cases} \frac{1}{2\eta}\|u\|_2^2 & \text{if } u \in S \\ 0 & \text{otherwise} \end{cases}$$ $\leftarrow$ S can be probability space

---

Regret with Entropy Regularizer

Assume: $f_1, f_2 \dots f_n$ are L-Lipschitz w.r.t. l_1-norm

$$\text{Regret}(n, u) \leq R(u) - \min_{u \in S} R(u) + \frac{L^2 n}{\sigma}$$

$$R(u) = \sum_{i=1}^{d} u_i \log u_i$$

$\dots \sum w_i = 1$

$u \in S = \{w, w \geq 0, \Sigma \ldots = 1$

$$= R(u) + \max_{u \in S} \left(-R(u)\right) + \frac{L^2 n}{\sigma}$$

$$-R(u) = \frac{1}{\eta} \sum_{i=1}^{d} u_i \log \frac{1}{u_i}$$ this is the entropy function

This is maximized when distribution is uniform

$$\leq \frac{1}{\eta} \log d \sum_{i=1}^{d} 1/d$$

$$\leq \frac{1}{\eta} \log d$$

$$= R(u) + \frac{1}{\eta} \log d + L^2 n \eta \quad [\sigma = 1]$$

$$\therefore \text{Regret}(n, u) \leq \frac{1}{\eta} \sum u_i \log u_i + \frac{1}{\eta} \log d + L^2 n \eta$$

Comparing with prediction with expert advice.

Regret in that case was: Expected loss in every round − loss with single best expert.

In order to bring in that here, consider u to be unit vectors

$$u \in \{e_1, e_2, \ldots e_d\}$$

$$e_1 = \{1, 0 \ldots 0\}$$
$$e_2 = \{0, 1, \ldots 0\}$$
$$\vdots$$
$$e_d = \{0, 0 \ldots 1\}$$

Now taking u to be any one of the e's

Then

$$\frac{1}{\eta} \sum u_i \log u_i = 0$$

$$\therefore \text{Regret}(n, u) \leq \frac{1}{\eta} \log d + L^2 n \eta$$

Optimizing over η $\quad \boxed{1/\eta^* = \sqrt{nL^2/\log d}}$

$$\text{Regret}(n, u) \leq \sqrt{nL^2 \log d} + \sqrt{nL^2 \log d}$$

$$\leq \sqrt{2}L\sqrt{2n\log d}$$

With weighted majority Regret $\leq \sqrt{2n\log d}$

with weighted majority : $f_t = \langle \omega_t, v_t \rangle$

$$|f_t(\omega) - f_t(u)| = |\langle \omega, v_t \rangle - \langle u, v_t \rangle|$$

$$= |\langle \omega - u, v_t \rangle|$$

$$\leq \|\omega - u\|_1 \|v_t\|_\infty$$ ← we need to find this value to get the Lipschitz const

V_t's are loss vectors hence

$\max V_t = 1$

Thus $f_t = \langle \omega_t, v_t \rangle$ is 1 - Lipschitz