Week 4 — Lecture 21

Regret bound of FoReL

We showed for FTL that

$$\sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(u) \leq \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(w_{t+1})$$

Lemma: Let $w_1, w_2 \ldots$ be a sequence of vectors produced by FoReL.

Then $\forall u \in S$

$$\sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(u) \leq R(u) - R(w_1) + \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(w_{t+1}) \quad \text{--- } (1)$$

FoReL Algo:

$$\underset{w}{\operatorname{argmin}} \sum_{i=1}^{t} f_i(w) + \underbrace{R(w)}_{\text{Let this be } f_0(w)} \quad \left[f_t(w) \text{ at } t=0 \right. \quad \text{--- } (2)$$

So then eq (1) becomes:

$$\sum_{t=0}^{n} f_t(w_t) - \sum_{t=0}^{n} f_t(u) \leq \sum_{t=0}^{n} f_t(w_t) - \sum_{t=0}^{n} f_t(w_{t+1})$$

$$\Rightarrow \cancel{f_0(w_0)} - f_0(u) + \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(u) \leq \cancel{f_0(w_0)} - f_0(w_1) + \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(w_{t+1})$$

$$\Rightarrow \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(u) \leq f_0(u) - f_0(w_1) + \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(w_{t+1})$$

$$\left.\begin{aligned} f_0(u) &= R(u) \\ \& \ f_0(w_1) &= R(w_1) \end{aligned}\right\} \text{from (2) above}$$

Thus we get back eq (1)

Now in order to work out the regret bound, let

$R(w) = \frac{1}{2\eta} ||w||^2$ and for $f_t(w) = \langle w, z_t \rangle$

also let $S \in \mathbb{R}^d$

Thus,

$$R(n, \text{FoReL}, u) \leq R(u) - \cancel{R(w_1)}^{0} + \sum_{t=1}^{n} f_t(w_t) - \sum_{t=1}^{n} f_t(w$$

$$= \frac{1}{2\eta} ||u||^2 + \sum_{t=1}^{n} \langle w_t \, z_t \rangle - \sum_{t=1}^{n} \langle w_{t+1} \, z_t \rangle$$

$$= \frac{1}{2\eta} ||u||^2 + \sum_{t=1}^{n} \langle w_t - w_{t+1}, z_t \rangle$$

$$= \frac{1}{2\eta} ||u||^2 + \sum_{t=1}^{n} \eta ||z_t||_2^2$$

FoReL:

$w_1 = \underset{w \in S}{\text{argmin}} \cancel{\sum_{i=1}^{t-1} f_i(w)}^{0} + R(w_1)$

Then

$w_1 = \underset{w}{\text{argmin}} ||w||_2^2$

Update rule:

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

Thus:

$$R(n, \text{FoReL}, u) \leq \frac{1}{} ||u||^2 + \eta \sum_{t=1}^{n} ||z_t||_2^2$$

2η

Let $\|Z_t\|_2^2 \leq L \;\; \forall t$; $U = \{ U : \|U\|_2^2 = B \}$

Then :

$$R(n, \text{FoReL}, u) \leq \frac{1}{2\eta} B + \eta \sum_{t=1}^{n} L$$

$$= \frac{B}{2\eta} + Ln\eta \quad \leftarrow \text{Convex function in } \eta$$

Now in order to choose an appropriate η, we need to differentiate twice, find out the slope w.r.t η and then choose the appropriate η

Hence $\eta = \sqrt{\frac{B}{2Ln}}$

$\therefore \; R(n, \text{FoReL}, u) \leq \sqrt{2BLn} \quad \leftarrow$ substituting η

$\therefore \; R(n, \text{FoReL}, u) \leq O\sqrt{n} \quad \leftarrow$ Thus sub-linear in n

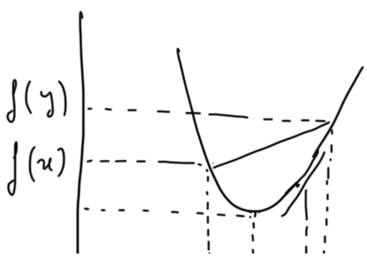
---

Linearization of convex functions

Def^n of a convex function:

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \text{--- } (3)$$

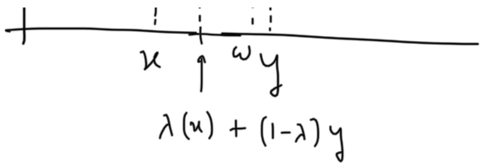
where $\lambda \in [0,1]$



at $\lambda = 0$

$f(\lambda x + (1-\lambda) y) = f(y)$

at $\lambda = 1$



$f(\lambda x + (1-\lambda)y) = f(x)$

Thus what (3) says is that $f(\lambda(x) + (1-\lambda)y)$ will lie within $f(x)$ & $f(y)$ $[\lambda \in [0,1]]$

Also at any point w, we can draw a tangent, which would be a lower bound on the convex function

<u>Lemma</u>: Let S be a convex set. A function

$$f : S \to \mathbb{R} \text{ is convex}$$

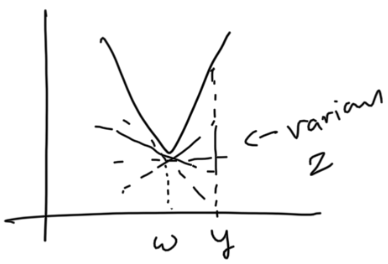
iff $\forall\, w \in S \;\exists\, z$ s.t.

$$\forall u \quad f(u) \geq f(w) + \langle u - w, z \rangle \qquad \text{— (4)}$$

This z is called sub-gradient of f at w.

Denoted as $\partial f(w)$

→ If f is differentiable and $\partial f(w)$ is a singleton then it is denoted as $\nabla f(w)$



At w, function is not differentiable. We can have various z satisfying the Lemma, hence at w we have a sub-gradient set.

However at point y, it is differentiable hence we have only one z and that's the gradient

Lemma (4) can be used to linearize any convex functio

$$R_n = \sum_{t=1} f_t(\omega_t) - \sum_{t=1} f_t(u)$$

$$\leq \langle \omega_t - u, z_t \rangle$$

$$= \langle \omega_t, z_t \rangle - \langle u, z_t \rangle$$

If function is convex & differentiable everywhere, ~~Assu~~ $\therefore z_t = \nabla f(\omega$

Thus any ~~functi~~ convex function can be linearized. Then the bound for linear function derived before can be utilized for any convex function.

Hence with FoReL we can use the Online Gradient Descent algorithm. (with L2 Regularizer)

Online Gradient Descent

Parameter $\eta > 0$

initialize $\omega_t = 0$

update rule $\omega_{t+1} = \omega_t - \eta z_t$ where $z_t \in \partial f_t(\omega_t)$