Week 4 - Lecture 22

For a loss function of the form:

$$f_t(\omega) = \langle \omega, z_t \rangle$$

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

we have the Online Gradient Descent algo for updates

OGD:

parameter: $\eta > 0$

initialize: $\omega_1 = 0$

update rule: $\omega_{t+1} = \omega_t - \eta z_t$

$$z_t \in \partial f(\omega_t)$$

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

where $\|u\|_2^2 \leq B \qquad \|z_t\|_2^2 \leq L$

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

$$R(n, u) \leq \sqrt{2BLn}$$

Regret is dependent on the size of the gradients, hence to ~~keep~~ control the gradients we set some constraints.

Gradients can also be thought of in terms of <u>Lipschitz functions</u>

f is L-Lipschitz w.r.t. norm $\|\cdot\|$ ← operator

$$|f(x) - f(y)| \leq L\|x-y\| \quad \forall\, x, y$$
— (2)

[x & y could be any vectors
L is constant]

<u>Norms</u> :

$$\|x\|_p = \left(\Sigma|x_i|^p\right)^{1/p} \quad [\text{l-p norm}]$$

If $p=1$, l-1 norm $\|x\|_1 = \Sigma|x_i|$

$p=2$, l-2 norm $\|x\|_2 = \sqrt{\Sigma x_i^2}$

A generic norm of x is denoted as $\|x\|$

p is not specified, it can be specified as anything.

<u>Dual norm</u> (generic def^n)

$$\|x\|_* = \max\{\langle w, x\rangle : \|w\| \leq 1\}$$

<u>Verify</u> :

$\langle w, z\rangle \leq \|w\|\,\|z\|_*$ [w & z can be interchanged

If $p, q \geq 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$ —— (3)

then l-p & l-q norms are dual.

Case 1

Let $p = 1$, then to satisfy (3) q ~~to~~ has to be ∞

$\therefore$ l-1 and l-∞ are dual.

l-∞ norm

$$\|x\|_\infty = \max_i |x_i|$$

Case 2

Let $p = 2$ then $q = 2$ to satisfy (3) $\therefore$ l-2 & l-2 are du

~~J~~ Thus l-2 norm is dual of itself.

The L in (2) can change depending on the norm.

Lemma: Let $f : S \longrightarrow R$ be convex

Then f is L-Lipschitz with respect to a norm $\|\cdot\|$

iff $\forall\, w \in S$ and $z \in \partial f(w)$ we have

$\|z\|_* \leq L$ where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$

What this essentially means is that if the function is L-Lipschitz then the sub-gradients are upper bounded by L.

Thus for eq 1 to be satisfied, one need not worry if sub-gradients are bounded by L, as long as the loss function is L-Lipschitz.

However in 1 $\|z_t\|_2^2 \leq L$ was bounded and in the above lemma it is $\|z\|_* \leq L$. Hence while computing the regret bounds henceforth one has to keep in mind the $\sqrt{L}$ instead of L, for the regret bound to hold.

Eq 1 holds for a specific regularizer

$$R(u) = \frac{1}{2\eta}\|u\|_2^2 \quad \text{—— } (4)$$

$R(u) = f_0(w)$ was assume (Lec 21

We need to verify that 4 is L-Lipschitz

& Substituting $R(u)$ in eq 2

$$|f(x) - f(y)| \leq L\|x-y\| \quad \text{—— } (2)$$

L.H.S:

$$\left|\frac{1}{2\eta}\|x\|_2^2 - \frac{1}{2\eta}\|y\|_2^2\right| = \frac{1}{2\eta}\left|\|x\|_2^2 - \|y\|_2^2\right|$$

$$= \frac{1}{2\eta}\left|\Sigma(x_i^2 - y_i^2)\right|$$

R.H.S:

$$\|x-y\|_2 = \Sigma(x_i - y_i)^2$$

<u>Prove</u>: L.H.S $\leq$ R.H.S

[not shown, left for students
to complete

Then $R(u)$ is L-Lipschitz with $L = \frac{1}{2\eta}$