CS 7545: Machine Learning Theory **Fall 2018**

Lecture 13: Follow The Regularized Leader

Lecturer: Jacob Abernethy *Scribes: Teng Li, Yeojoon Youn*

13.1 Review

Last time we introduced follow the leader algorithm:

Algorithm 1: Follow the leader

Let K be a convex subset of R^d
For t = 1...T
  Alg choose $x_t \in K$
  Nature Choose $f_t : K \to R$
  $Regret^u_T = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(u)$
End for

The algorithm choose $x_{t+1} = argmin_{x\in K} \sum_{s=1}^t f_s(x)$, and u is some point in K.

Last class we discussed but didn't prove following fact:
If f_t is α_t-strongly convex with respect to sore norm, then $Reg_T(FTL) = O(\sum_{t=1}^T \frac{1}{A_t})$ *where* $A_t = \sum_{s=1}^t \alpha_s$.
If $\forall t, \alpha_t = 1$, we can get $Reg_T(FTL) = O(logT)$

13.2 Solve Linear Case

It's hardest case. Every online convex optimization problem can reduce to linear loss functions.

Trick: Given a sequence of $f_1...f_T$
 Algorithm chooses $x_1...x_T$
 Define $\tilde{f}_t(x) := \langle \nabla_{x_t} f_t, x - x_t\rangle + f_t(x_t)$
 $\tilde{f}_t(x_t) = f_t(x_t)$ for t=1...T
 $\tilde{f}_t(u) \leq f_t(u)\ \forall u \in K$
 $\forall u \in K, \sum_{t=1}^T (f_t(x_t) - f_t(u)) \leq \sum_{t=1}^T (\tilde{f}_t(x_t) - \tilde{f}_t(u))$
 Hence it's "harder" to deal with linear loss

Definition 13.1 $f_t(x) = g_t \cdot x$ *(g_t is the gradient of f_t)*
Observation: FTL in linear case (Online Linear Optimization) has bad regret(linear in T).

Example: Given $K = [-1, 1], x_t \in K$

$$f_t(x_t) = \begin{cases} \frac{1}{2}x, \ t = 1 \\ x, \ t > 1 \ and \ t \ is \ odd \\ -x, \ t > 1 \ and \ t \ is \ even \end{cases}$$

FTL: $x_t = argmin_{x \in [-1,1]} \sum_{s=1}^{t-1} f_s(x) = \begin{cases} -1, \ t = 2 \\ 1, \ t = 3 \\ -1, \ t = 4 \end{cases}$

$\sum_{t=2}^{T} f_t(x_t) = T - 1$
$-1 \leq \sum_{t=1}^{T} f_t(u) \leq 1$
Then we get Regret$\geq T - 2$

13.3 Follow the Regularized Leader(FTRL)

Let R be any convex function on domain.
Also, let's consider the FTRL in linear case. ($f_t(x) = g_t \cdot x$)

Algorithm 2: Follow the regularized leader

Let K be a convex subset of R^d
For t = 1...T
  Alg choose $x_t \in K$
  Nature Choose $f_t : K \to R$
  $x_{t+1} = \underset{x \in K}{\text{argmin}} \ \eta \sum_{s=1}^{t} f_s(x) + R(x)$
  $\quad = \underset{x \in \mathbb{R}^d}{\text{argmin}} \ \eta \sum_{s=1}^{t} g_s \cdot x + R(x)$
  $Regret_T^u = \sum_{t=1}^{T} g_t \cdot x_t - \sum_{t=1}^{T} g_t \cdot u$
End for

Lemma 13.2

$$\eta g_t \cdot (x_t - u) = D_R(u, x_t) - D_R(u, x_{t+1}) + D_R(x_t, x_{t+1})$$

Proof: Since $x_{t+1} = \underset{x \in \mathbb{R}^d}{\text{argmin}} \ \eta \sum_{s=1}^{t} g_s \cdot x + R(x)$, gradient of $\eta \sum_{s=1}^{t} g_s \cdot x + R(x)$ at x_{t+1} should be 0.
Let's apply this observation to x_{t+1}, x_t.

$$x_{t+1} : \eta \sum_{s=1}^{t} g_s + \nabla R(x_{t+1}) = 0$$
$$x_t : \eta \sum_{s=1}^{t-1} g_s + \nabla R(x_t) = 0$$
$$\Rightarrow \nabla R(x_t) - \nabla R(x_{t+1}) = \eta g_t$$

Therefore,

$$\begin{aligned} D_R(u,x_t) - D_R(u,x_{t+1}) + D_R(x_t,x_{t+1}) &= R(u) - R(x_t) - \nabla R(x_t)\cdot(u-x_t) \\ &\quad - (R(u) - R(x_{t+1}) - \nabla R(x_{t+1})\cdot(u-x_{t+1})) \\ &\quad + R(x_t) - R(x_{t+1}) - \nabla R(x_{t+1})\cdot(x_t - x_{t+1}) \\ &= (\nabla R(x_t) - \nabla R(x_{t+1}))\cdot(x_t - u) \\ &= \eta g_t \cdot (x_t - u) \end{aligned}$$

■

Lemma 13.3

$$\eta \sum_{t=1}^{T} g_t \cdot (x_t - u) = D_R(u,x_1) - D_R(u,x_{T+1}) + \sum_{t=1}^{T} D_R(x_t,x_{t+1})$$

Proof: Applying telescoping to Lemma 13.2. ■

Theorem 13.4 (Regret Upper Bound of FTRL) *If R is 1-strongly convex with respect to $\|\cdot\|$, then*

$$\mathrm{Reg}_T^u(\mathrm{FTRL}) \le \frac{1}{\eta}(R(u) - \min_{x\in\mathbb{R}^d} R(x)) + \sum_{t=1}^{T} \eta \|g_t\|_*^2$$

Proof: Let's consider $D_R(u,x_1), D_R(x_t,x_{t+1})$.

$$\begin{aligned} D_R(u,x_1) &= R(u) - R(x_1) - \nabla R(x_1)(u-x_1) \ (\because \nabla R(x_1) = 0 \ \Leftarrow x_1 : \text{ minimizing the R}) \\ &= R(u) - \min_{x\in\mathbb{R}^d} R(x) \end{aligned}$$

$$\begin{aligned} D_R(x_t,x_{t+1}) &\le D_R(x_t,x_{t+1}) + D_R(x_{t+1},x_t) \ (\ge \|x_t - x_{t+1}\|^2 \ \Leftarrow \ \because R: \text{ 1-strongly convex}) \\ &= R(x_t) - R(x_{t+1}) - \nabla R(x_{t+1})\cdot(x_t - x_{t+1}) \\ &\quad + R(x_{t+1}) - R(x_t) - \nabla R(x_t)\cdot(x_{t+1} - x_t) \\ &= (\nabla R(x_t) - \nabla R(x_{t+1}))\cdot(x_t - x_{t+1}) \\ &= \eta g_t\cdot(x_t - x_{t+1}) \\ &\le \eta\|x_t - x_{t+1}\|\|g_t\|_* \end{aligned}$$

Note that $\|x_t - x_{t+1}\|^2 \le \eta\|x_t - x_{t+1}\|\|g_t\|_* \ \Rightarrow \ \|x_t - x_{t+1}\| \le \eta\|g_t\|_*$.

$$\begin{aligned} \therefore D_R(x_t,x_{t+1}) &\le \eta\|x_t - x_{t+1}\|\|g_t\|_* \\ &\le \eta^2\|g_t\|_*^2 \end{aligned}$$

Finally,

$$\begin{aligned} \mathrm{Reg}_T^u(\mathrm{FTRL}) &= \sum_{t=1}^{T} g_t\cdot(x_t - u) \\ &= \frac{1}{\eta}D_R(u,x_1) - \frac{1}{\eta}D_R(u,x_{T+1}) + \frac{1}{\eta}\sum_{t=1}^{T} D_R(x_t,x_{t+1}) \\ &\le \frac{1}{\eta}(R(u) - \min_{x\in\mathbb{R}^d} R(x)) + \sum_{t=1}^{T} \eta\|g_t\|_*^2 \end{aligned}$$

■

Furthermore, assuming that $R(u) - \min_{x \in \mathbb{R}^d} R(x) \leq R_{\max},\ \|g_t\|_* \leq G$

$$\text{Reg}_T^u(\text{FTRL}) \leq \frac{1}{\eta} R_{\max} + TG^2\eta$$

Tuning η ($\eta = \frac{1}{G}\sqrt{\frac{R_{\max}}{T}}$)

$$\therefore \text{Reg}_T^u(\text{FTRL}) \leq 2G\sqrt{R_{\max}T}$$