EECS598: Prediction and Learning: It's Only a Game Fall 2013

Lecture 21: Bandit Algorithm / Blackwell Approachability

Prof. Jacob Abernethy *Scribe: Tae Hyung Kim*

Announcements

- HW3 is due Nov 27 (next Wednesday)
- Work on projects!! (presentation after three weeks)

21.1 Bandit problem in stochastic shortest path (continue on the last lecture)

21.1.1 FTRL in the bandit setting

In every round,

$$x_t = \arg\min_{x\in K} \sum_{s=1}^{t-1} f_s \cdot x + \lambda R(x) \tag{21.1}$$

From the last lecture, we can use estimated loss function. Therefore,

$$x_t = \arg\min_{x\in K} \sum_{s=1}^{t-1} \tilde{f}_s \cdot x + \lambda R(x) \tag{21.2}$$

The regret bound becomes,

$$\begin{aligned} Regret_T &\le \sum_{t=1}^{T} \lambda D_R(x_t, x_t+1) + \lambda R(x^\star) \\ &\le \sum_{t=1}^{T} \frac{\|\tilde{f}_t\|_\star^2}{\lambda} + \lambda R(x^\star) \\ &\le \frac{T\cdot G}{\lambda} + \lambda D \\ &\le 2\sqrt{T\cdot G\cdot D} \end{aligned} \tag{21.3}$$

Probelm As x_t approaches to the boundary, $\|\tilde{f}_t\|_\star^2$ grows very large. So above inequality breaks.

Solution: Use regularization with Self Concordance Function

21.1.2 Self Concordance Function

Classical Newton's method Let our objective function be $g(x)$. we want to minimize

$$\min_{x\in D} g(x) \tag{21.4}$$

By adding self-concordance regularization term R,

$$\min_{x \in D} g(x) + \lambda R(x) \tag{21.5}$$

The Newton's update rule becomes

$$x_{t+1} \leftarrow x_t + (\nabla^2_{x_t} R)^{-1} \nabla \hat{g}(x_t) \tag{21.6}$$

Therefore, x_{t+1} is in the ellipsoid centered on x_t.

$$x_{t+1} \in (\nabla^2_{x_t} R)\text{-ellipsoid}$$

Back to Bandit Optimization Previously, our update rule was

$$x_t = \arg\min_{x \in K} \sum_{s=1}^{t-1} \tilde{f}_s \cdot x + \lambda R(x) \tag{21.7}$$

We approximate f_t as eigenpoles of $(\nabla^2_{x_t} R)$-elliposid.

$$\tilde{f}_t \approx \lambda_i^{1/2} e_i \tag{21.8}$$

where λ_j and e_j are eigenvalues and unit eigenvalues of $\nabla^2_{x_t} R$

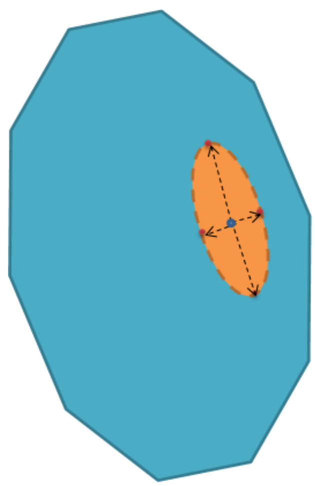

Figure 1: eigenpols approximation

Then, the new regret bound becomes

$$\begin{aligned} Regret_T &\leq \sum_{t=1}^{T} \lambda D_R(x_t, x_t + 1) + \lambda R(x^\star) \\ &\leq \sum_{t=1}^{T} \frac{\tilde{f}_t^T (\nabla^{-2}_{x_t} R) \tilde{f}^T}{\lambda} + \lambda R(x^\star) \\ &\leq \frac{nG\sqrt{\sigma_{i_t}} \sigma_{i_t}^{-1} \sqrt{\sigma_{i_t}}}{\lambda} + \lambda D\theta logT \\ &\leq 2\sqrt{n \cdot G \cdot D \cdot \theta \cdot TlogT} \end{aligned} \tag{21.9}$$

However, this results are "in expectation" only, and only work against "oblivious adversaries". The general problem is still hard.

21.2 Blackwell Approachability

In standard 2-player 0-sum game, the game matrix M satisfies

- $M \in [0,1]^{n\times m}$,
- $M_{ij} \in \mathbb{R}^1$ is the payoff for P1, when P1 and P2 play i and j respectively.

The minimax theorem is

$$\min_p \max_q p^T M q = \max_q \min_p p^T M q \tag{21.10}$$

or equivalently, (strong duality)

$$\begin{aligned} \forall p \exists q : p^T M q \geq c \\ \exists q \forall p : p^T M q \geq c \end{aligned} \tag{21.11}$$

Generation Now, we want to generalize this to the case when $M_{ij} \in \mathbb{R}^d$.

Let $r(i,j)$ be the payoff vector for P1,

$$r : \Delta_n \times \Delta_m \to \mathbb{R}^d$$

This is biaffine!

1. $r(\alpha p_1 + (1-\alpha)p_2, q) = \alpha r(p_1, q) + (1-\alpha) r(p_2, q)$
2. $r(p, \alpha q_1 + (1-\alpha)q_2) = \alpha r(p, q_1) + (1-\alpha) r(p, q_2)$

In this generalized version, we can define the minimax theorem by

$$\begin{aligned} \forall p \exists q : r(p,q) \in S \\ \exists q \forall p : r(p,q) \in S \end{aligned} \tag{21.12}$$

S is a certain convex set applying some constraints. In general, this condition is not satisfied.

Example: A bad case

$$\begin{gathered} r(p,q) = (p,q) \\ S = \{(x,y) : x = y\} \end{gathered}$$

$\forall p$, there exists $q = p$ such that $r(p,q) \in S$.
However, there is no q satisfying $\forall p : r(p,q) \in S$

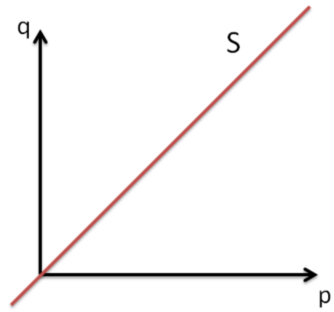

Figure 2: A bad case

21.2.1 Blackwell Approachability Theorem

If r, S, p, q satisfies

$$\forall p, \exists q : r(p,q) \in S \tag{21.13}$$

Then, $\exists$ an adaptive strategy

$$q_t \leftarrow f(p_1, p_2, ..., p_t) \tag{21.14}$$

such that

$$\frac{1}{T}\sum r(p_t, q_t) \rightarrow S \tag{21.15}$$

or equivalently,

$$dist(\frac{1}{T}\sum r(p_t, q_t), S) \rightarrow 0 \tag{21.16}$$

Example In above bad case example, one possible strategy for q is to choose previous p.

$$q_t \leftarrow p_{t-1} \tag{21.17}$$

This strategy satisfies Blackwell Approachability theorem,

$$dist(\frac{1}{T}\sum_{t=1}^{T} p_t, \frac{1}{T}\sum_{t=0}^{T-1} p_t) \rightarrow S \tag{21.18}$$

where $p_0 = q_1$. (initial choice of q)

21.2.2 Halfspace condition

$\forall$ halfspaces $H \supset S$, $\exists q$ $\forall p$

$$r(p,q) \in H \tag{21.19}$$

Lemma: The followings are equivalent

1. The halfspace condition
2. $\forall$ halfspaces $H \supset S$, $\forall p$, $\exists q$: $r(p,q) \in H$
3. $\forall p$, $\exists q$: $r(p,q) \in S$

Proof:

1. 1 and 2 are equivalent
   Project $r(p,q)$ into the normal of H, and apply minimax theorem.

2. $3 \Rightarrow 2$
   Assume $\exists H, \exists p_{bad}, \forall q$

$$r(p_{bad}, q) \notin H$$

   which implies $r(p_{bad}, q) \notin S$ (contradiction)

3. $1 \Rightarrow 3$
   If $\exists p_{bad}, \forall q$: $r(p_{bad}, q) \notin S$
   $\Rightarrow \exists$ hyperplane separating S and $\{r(p_{bad}, q) : q \in \Delta_m\}$, but this hyperplane violates 1.
   (contradiction)

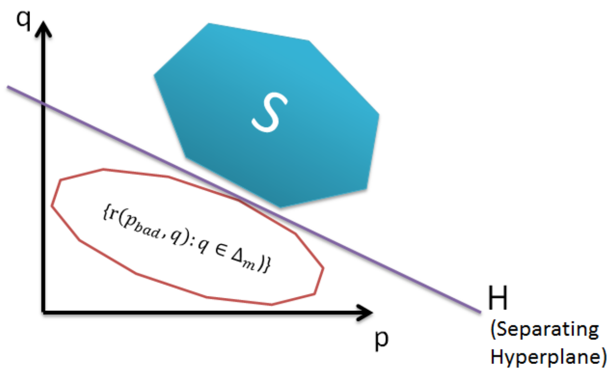

Figure 3: Separating Hyperplane