ALG-ML SEMINAR

LECTURE #6

OCTOBER 8, 2015

LECTURER: SÉBASTIEN BUBECK SCRIBE: KIRAN VODRAHALLI

Contents

1 Overview

Use the Bayesian game setting and apply information theory to bound the changes in entropy. Then apply this to regret by bounding the expectation of the sum of the losses by this information inequality and apply Cauchy-Schwarz. Then use a lemma on ability to distinguish between convex functions randomly sampled (via posterior distribution) over a convex space $\mathcal{K}$ to let you use pidgeon-hole to find a $\hat{x}$ for use in the ϵ-greedy Thompson sampling to pick an optimal strategy.

2 Convex Bandit Setting

We have $t = 1, 2, \cdots, T$ and $\mathcal{K} \subset \mathbb{R}^n$ a convex body. The player chooses some $x_t \in \mathcal{K}$. Then the adversary chooses some $l_t : \mathcal{K} \to [0, 1]$ where l_t is convex and 1-Lipschitz. The player then suffers some loss $l_t(x_t)$ and only sees the loss for their play. We define performance by looking at regret $R_T = \sum_{t=1}^T l_t(x_t) - \min_{x\in\mathcal{K}} \sum_{t=1}^T l_t(x)$. We can define minimax regret as $\mathcal{R}_T = \inf_{player} \sup_{adversary} \mathbf{E}[R_T]$. You can think of this as choosing the best thing in advance.

3 History

Kleinberg $'04$ got $\mathcal{R}_T \leq n^3 T^{3/4}$, and FKM $'04$ got $\mathcal{R}_T \leq \sqrt{n} T^{3/4}$. In 2015, we have the best bound. The strategy to get these results is just the stochastic gradient descent strategy: $x_{t+1} = \prod_{\mathcal{K}} (x_t - \eta \nabla l_t(x_t))$ from Robbins-Monroe in 1951. Here, nature does not provide the randomness, you do.

The only thing we want know is that we are going to replace ∇l_t with g_t, where $\mathbf{E}[g_t] \approx \mathbf{E}[\nabla l_t(x_t)]$. We can use the divergence theorem to get $g_t = (n/\delta) l_t(x_t + \delta u) u$ where u is uniform on S^{n-1}. The expectation $\mathbf{E}_u[g_t] = \nabla \bar{l}_t(x_t)$ where $\bar{l}_t(x) = \mathbf{E}_{b\in \text{ball}}[l_t(x + \delta b)]$. This is just a smoothing operation. Now it is really simple, you use the usual bound of gradient descent. You need to control the norm $\|g_1\|_2^2 \leq (n/\delta)^2$. This is still the best bound in some sense as it is polynomial time. After this there has been a long line of work trying to improve this under various conditions. If you assume loss functions are smooth, you can get $\mathcal{R}_T^{smooth} \leq T^{2/3}$. If you are strongly convex, you get this same bound. Then recently Elad Hazan proved $\mathcal{R}_T^{smoothandstronglyconvex} \leq \sqrt{T}$. But this is all the same algorithm roughly - the core idea is the same. You add a little bit because you need to play with certain thing. Linear functions are important though: For linear loss functions, then we know much more. We have $\min_{\mathcal{K}} \mathcal{R}_T^{linear} = \Theta(n\sqrt{T})$. Most of the work has been going in this direction. What was open was general convex functions. But most of the work focused on the linear case. You can formulate all of these problems like shortest path etc, which are linear in their representation. What is so special about linear? The gradient is the same everywhere, so you can get an extremely strong estimator for the gradient everywhere. We can now say that we understand the linear case very well. Not completely: We still don't have complete understanding of how the set $\mathcal{K}$ comes in: You can maybe do better for specific $\mathcal{K}$, but the bound is tight for the worst case.

4 New stuff: Convex Bandits

I will now tell you why I am interested. How can you squeeze as much as possible out of limited information. We didn't know whether you can get $\sqrt{T}$ regret in the general case. The theorem I am going to tell you is that you can. Even in dimension 1, this was open.

Theorem 4.1. *Bubeck, Dekel, Kren, Peres* $'15$ *for* $n = 1$, *Bubeck and Eldan* $'15$ *(full generality).*

$$\mathcal{R}_T \leq n^{11}\sqrt{T}$$

For the stochastic case, this also works.

Proof.

1. Step 1. $\mathcal{R}_T = \text{sup}_{adversary,distribution\mu on(l_1,\cdots,l_T)}\text{inf}_{player}\mathbf{E}[\mathcal{R}_T]$ by the minimax theorem which you can apply because we assume the oblivious case. The adversary is choosing a distribution of the loss sequences, and the player is choosing as before. The player is playing after the adversary. Now we can adapt to the distribution. This is a Bayesian game. Now we are going to bound this: We are going to provide a Bayesian strategy with $\sqrt{T}$ regret: But this doesn't get you a strategy for the primal problem. This is going to be a computational strategy with T^T after discretizing all the solutions (losses, etc.) - then just find it (so this is still computable).

2. Step 2. We are going to apply Russo and Van Roy information ratio: We define $n_t(x)$ is expected instantaneous regret at time t to x. Let $x^* \in \text{argmin}_x \sum_{t=1}^T l_t(x)$. To be clear, consider this a random variable: We are going to use some prior distribution of the past. Thus we can have a posterior distribution given everything we have seen so far. Then

   $$n_t(x) = \mathbf{E}_{ton(x_s,l_s(x_s))_{s<t}}[l_t(x) - l_t(x^*)] = \mathbf{E}_t[l_t(x)|x^*]$$

   which is for fixed x, thus only randomness is in adversary. We are not really taking expectation over t, the subscript t denotes a posterior distribution over t. We can write $x^*|(x_s, l_s(x_s))_{s<t}$. We can write this posterior just fine. This is the expectation of what we can expect to gain after playing x. There could be some variability in that $l_t(x)$ which is irrelevant to x^*. Thus we are going to average out all this noise and just look at what we get from x^*. x^* is a random variable because adversary has not revealed its l_t. Even at T, x^* is a random variable. Then we look at the conditional variance

   $$v_t(x) = \text{Var}_t\left[\mathbf{E}_t[l_t(x)|x^*]\right]$$

   The inner expectation is a variable of x^*, and the variance is taken over x^*. Let us re-write this quantity. The following lemma is amazing

   Lemma 4.2. *Russo and Roy* $'14$.

   $$\boldsymbol{E}[\sum_{i=1}^T v_t(x_t)] \leq \frac{1}{2}H(x^*)$$

If we have an ϵ-net over $\mathcal{K}$, then $n^* = \min_{\epsilon-net} H(x^*) \leq \log|\epsilon - net| \approx n\log(1/\epsilon)$.

Proof. Let us introduce $\alpha(t) = \mathbf{P}_t\{(x^* = x)\}$. Then $v_t(x) = \sum_z \alpha_t(z)\left(\mathbf{E}_t[l_t(x)|x^* = z] - \mathbf{E}_t[l_t(x)]\right)^2 \leq \frac{1}{2}\sum_z \alpha_t(z)\text{Entropy}\left(L_t(l_t(x)|x^* = z)||L_t(l_t(x))\right) = \mathcal{I}_t(x^*, l_t(x))$, recognizing the mutual information. You can view mutual information between x and y as how much decreasing entropy you get. We have $\frac{1}{2}\mathcal{I}_t(x^*, l_t(x)) = \frac{1}{2}\left(H_t(x^*) - H_t(x^*|l_t(x))\right)$. And now the magic happens: This gives $\mathbf{E}[v_t] \leq H_{t+1}(x^*) - H_t(x^*)$ which gives us the required telescoping sum. □

So this is a kind of information theoretic view of optimization. The adversary has some prior, I will adapt some to this prior, this is a Bayesian game, how do I analyze it? Information theory. All of these results are completely universal: Any Bayesian game, how does entropy progress. No convexity or bandits appeared. So two things left: I need to tell you how to use this for convexity and bandits.

3. Assume that we have a strategy such that $\mathbf{E}_t[n_t(x_t)] \leq \sqrt{\mathbf{E}_t[v_t(x_t)]}$. Then you can just sum these things (which will be your regret) and apply Cauchy-Schwartz. If you have this, then your expected regret is bounded by $\sqrt{(T/2)H(x^*)} = \sqrt{(T/2)\log K}$, where K is the size of the ϵ-net (you don't really need to discretize, it just simplifies the discussion). If you have this, then you are done. So the only question is can prove your inequality. Russo showed that Thompson sampling $\mathbf{P}_t\{x_t = z\} = \alpha_t(z)$ for any z, called "probability matching". This is the first bandit strategy - more general than multiarmed bandit strategy. The probability you play z is equal to the posterior distribution on z (the probability that it is optimal). Then $\mathbf{E}_t[n_t(x_t)] \leq \sqrt{K\mathbf{E}_t[v_t x_t]}$. Then the new idea is to not use Thompson sampling. This is ϵ-greedy meets Thompson sampling. $x_t = \text{argmin} n_t(x)$ the best point with probability $1-\epsilon$ or some $\hat{x}$ with probability ϵ. We can tune this ϵ by tuning the posterior. Now how do you pick this exploration point? First we show that ϵ is good, and the next twenty pages are about $\hat{x}$. How do you select it? Here is where convex geometry comes into play (this is also general, and was not previously known):

   Theorem 4.3. *Let $f : \mathcal{K} \to \mathbb{R}^+$ convex and 1-Lipschitz. Fix $\epsilon > 0$. Then there exists a distribution μ on $\mathcal{K}$ such that for all $g, \mathcal{K} \to \mathbb{R}$, convex, 1-Lipschitz s.t. exists $z \in \mathcal{K}$ with $g(z) \leq -\epsilon$ such that $\mu(\{x : |f(x) - g(x)| \geq \epsilon(1/n^3)\}) > \frac{1}{n^7\log(1/\epsilon)}$. You can view this as a statement about hypothesis testing. Thus we can differentiate between the two. You do not only care about minimizing f, you care about exploring f also. Remember that this is the Bayesian setting: f is what we have seen so far. f is the conditional expectation of the loss given the past. What you are trying to do is find a good sigma algebra (subspace on which you are projecting, enhance your posterior distribution, make your posterior as informative as possible, which is orthogonal to being at any given time step) also. The proof is constructive, but at some point we use the probabilistic method. This basically says that a convex function must be mostly everywhere smooth (like in Rademacher's theorem, use Poincare's inequality: Use small balls to cut off parts of the space - multi-scale argument). The adversary can try to hide in the non-smoothness of the function, in the end it is much more by hand.*

   Once you have μ, you can use pidgeon-hole to find a good $\hat{x}$. You should think of g as the expectation of the loss - the key point of this theorem is can we tell the difference betweeen what I can see and what I expect. Convexity allows you to move from ϵ^n to the log term. You can't sample from μ efficiently. □