Homework 5

Akshay Krishnamurthy
Due: Thursday 11/16

November 12, 2017

Instructions: Turn in your homework in class on Thursday 11/16/2017

1. **Log-barrier regularization.** The Agile Mirror Descent algorithm is the following. Starting with w_1, at round t

   (a) Play w_t, observe loss f_t, let $\ell_t = \nabla f_t(w_t)$.

   (b) Let $\tilde{w}_{t+1}$ satisfy $\nabla R(\tilde{w}_{t+1}) = \nabla R(w_t) - \eta \ell_t$

   (c) set $w_{t+1} = \text{argmin}_{w \in S} D_R(w||\tilde{w}_{t+1})$

   In this problem we will analyze this algorithm. It will be helpful to use the generalized Pythagorean theorem: Let $x \in S, \tilde{y}$ by some vector, and define $y = \text{argmin}_{z \in S} D_R(z||\tilde{y})$, then $D_R(x||\tilde{y}) \geq D_R(x||y) + D_R(y||\tilde{y})$. (Bonus: prove the generalized pythagorean theorem)

   (a) Prove a Be-The-Leader style lemma for this algorithm. For any $w \in S$ and t

   $$\langle w_t - w, \ell_t \rangle \leq \langle w_t - \tilde{w}_{t+1}, \ell_t \rangle + \frac{1}{\eta} \left(D_R(w||w_t) - D_R(w||w_{t+1}) \right)$$

   (b) Now, in the Experts setting (i.e. $S = \Delta([d])$), using the log-barrier regularizer $R(w) = \sum_{i=1}^{d} - \log(w_i)$, derive a closed form for $\tilde{w}_{t+1}$ in terms of w_t.

   (c) Use the above two steps to prove the regret bound

   $$\text{Reg}_T(u) \leq \eta \sum_t \langle w_t^2, \ell_t^2 \rangle + \frac{1}{\eta}(R(u) - R(w_1))$$

   Note this regularizer has some very different properties than entropic regularizer that we saw in class and it can be quite useful in various contexts. For example, Foster et al. [4] for learning in game-theoretic settings and Agarwal et al. [1] use it for model selection. The useful thing is that the first term is $\langle w_t^2, \ell_t^2 \rangle$ instead of the usual $\langle w_t, \ell_t^2 \rangle$, which is important for settings with partial feedback.

2. **Online Gradient Descent.** In this problem we'll study an adaptive online gradient method with a time-varying step size. We use the learning rate $\eta_t = \frac{B}{\sqrt{\sum_{\tau=1}^{t} \|\nabla f_\tau(w_\tau)\|_2^2}}$ where the (convex) loss function at round τ is f_τ and we assume that $\|w\| \leq B$ for all w we are competing with. In more detail, the update is $w_{t+1} = w_t - \eta_t \nabla f_t(w_t)$. In this problem we will prove that for all times T, we have

   $$\text{Reg}(T) = \sum_{t=1}^{T} f_t(w_t) - \min_w \sum_{t=1}^{T} f_t(w) \leq 3B \sqrt{\sum_{t=1}^{T} \|\nabla f_t(w_t)\|_2^2}$$

   (a) First analyze online gradient descent with variable learning rate η_t at round t. Follow the gradient descent proof to show that

   $$\text{Reg}(T) \leq \frac{B^2}{2\eta_T} + \sum_{t=1}^{T} \eta_t \|\nabla f_t(w_t)\|_2^2$$

(b) Next, use induction to prove the analytical inequality

$$\sum_{t=1}^{T} \frac{a_t}{\sum_{i=1}^{t} a_i} \leq 2\sqrt{\sum_{t=1}^{T} a_t}$$

(c) Put the two parts together to obtain the regret bound.

This problem is about adaptive gradient methods which are now the default optimization algorithm in most packages. The main improvement is that we don't use any bound on the gradients to set the learning rate, so we can adapt to situations where the gradients are much smaller than the bound. Duchi et al. [3] analyze several other adaptive gradient methods.

3. **Stochastic Bandits with delays.** In the stochastic K-arm bandit setting, suppose now that the loss for round t is only revealed at round $t + \gamma$ for $\gamma \in \mathbb{N}$. Derive a instance-dependent regret bound for a confidence based algorithm.

$$\text{Reg}(T) = T\mu^\star - \sum_{t=1}^{T} \mu(a_t) \leq O\left(\sum_{a \neq a^\star} \frac{\log(T)}{\Delta_a} + \gamma \sum_{a \neq a^\star} \Delta_a\right)$$

where $\Delta_a = \mu(a^\star) - \mu(a)$ and $a^\star = \text{argmax}_a \, \mu$.

Handling delayed feedback arises naturally in many settings (see for example Cesa-Bianchi et al. [2] for other bandit problems). It is reassuring to know that delays are not really a major issue in most problems.

References

[1] Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire. Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pages 12–38, 2017.

[2] Nicolo Cesa-Bianchi, Claudio Gentile, Yishay Mansour, and Alberto Minora. Delay and cooperation in nonstochastic bandits. In *Conference on Learning Theory*, pages 605–622, 2016.

[3] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

[4] Dylan J Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Eva Tardos. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems*, pages 4734–4742, 2016.