COMS6998-11: Homework 1

Akshay Krishnamurthy
akshay@cs.umass.edu
Due: Thursday 3/3

Instructions: Turn in your homework to me by email by Monday 2/28.

1. **First order generalization.** In this problem we will prove a stronger generalization error bound for the agnostic binary classification, that uses more information about the distribution. Let P be a distribution over (X, Y) pairs where $X \in \mathcal{X}$ and $Y \in \{+1, -1\}$ and let $\mathcal{F} \subset \mathcal{X} \to \{+1, -1\}$ be a finite hypothesis class and let ℓ denote the zero-one loss $\ell(\hat{y}, y) = \mathbf{1}\{\hat{y} \neq y\}$. Let $R(f) = \mathbb{E}\ell(f(X), Y)$ denote the risk, and let $f^\star = \mathrm{argmin}_{f \in \mathcal{F}} R(f)$. Given n samples let $\hat{f}_n$ denote the empirical risk minimizer. The goal here is to prove a sample complexity bound of the form:

$$R(\hat{f}_n) - R(f^\star) \leq c_1 \sqrt{\frac{R(f^\star)\log(|\mathcal{F}|/\delta)}{n}} + c_2 \frac{\log(|\mathcal{F}|/\delta)}{n}. \tag{1}$$

for constants c_1, c_2. This can be a much better bound than the one we saw in class, if $R(f^\star)$ is small. In particular, if $R(f^\star) = 0$ (which is called the realizable setting), then this bound obtains a $1/n$-rate.

   (a) To prove the result, we will use Bernstein's inequality, which is a sharper concentration result.

   Theorem 1 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_n$ be iid real-valued random variables with mean zero, and such that $|X_i| \leq M$ for all i. Then for all $t > 0$*

$$\mathbb{P}[\sum_{i=1}^n X_i \geq t] \leq \exp\left(-\frac{t^2/2}{\sum_{i=1}^n \mathbb{E}[X_i^2] + Mt/3}\right).$$

   We will not prove this here. Use the inequality to show that with probability at least $1 - \delta$

$$|\bar{X}| \leq \sqrt{\frac{2\mathbb{E}[X_1^2]\log(2/\delta)}{n}} + \frac{2M\log(2/\delta)}{3n}. \tag{2}$$

   where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and X_is satisfy the conditions of Bernstein's inequality.

   (b) Use Eq. (2) and the union bound to show Eq. (1).

2. **Action elimination for MAB.** In class, we saw the UCB strategy for regret minimization in the stochastic MAB problem. Another strategy, known as *action elimination*, also has similar properties. Consider the stochastic A-armed bandit problem in the *reward* formulation, where there are distributions $\nu(a)$ for each arm $a \in [A]$, each with mean $\mu(a)$. We view the samples as rewards, so regret is measured as

$$\mathrm{Regret}_T = \max_a T \cdot \mu(a) - \sum_t \mu(a_t).$$

   Active arm elimination maintains a set of "active arms" which could plausibly be the best one and operates in phases or epochs. We start with $\mathcal{A}_1 = [A]$ and epoch $i = 1$. Then in epoch i we:

   (a) For each arm $a \in \mathcal{A}_i$ play a for 2^i rounds.

   (b) Update the empirical mean $\hat{\mu}(a)$ and confidence $\mathrm{conf}(a)$ accordingly

(c) Update $\mathcal{A}_{i+1} = \{a \in \mathcal{A}_i : \hat{\mu}(a) + \mathrm{conf}(a) \geq \max_{a'} \hat{\mu}(a') - \mathrm{conf}(a')\}$

In the last step, we keep arm a if, based on its optimistic estimate $\hat{\mu}(a) + \mathrm{conf}(a)$, it could still be the best arm. Show that this algorithm has a $\sqrt{AT\log(AT/\delta)}$ regret bound. You may assume that if arm a has been pulled N times, we can set $\mathrm{conf}(a) = \sqrt{\log(AT/\delta)/N}$.

3. **Action elimination for linear bandits.** It's also possible to design action elimination algorithms for the linear stochastic bandit problem but avoiding a dependence on the number of actions is tricky. Suppose we have a finite but large arm set $\mathcal{A}$ with features $\{x_a : a \in \mathcal{A}\} \subset \mathbb{R}^d$ where $\|x_a\|_2 \leq B \forall a \in \mathcal{A}$. We assume that $\mathbb{E}[r \mid a] = \langle x_a, \theta^\star\rangle$ for some unknown parameter $\theta^\star$ and that the rewards are bounded in, say $[0,1]$. We measure regret as

$$\mathrm{Regret}_T = \max_a T \cdot \langle x_a, \theta^\star\rangle - \sum_t \langle x_{a_t}, \theta^\star\rangle .$$

Here we will not use an epoching algorithm, but the high-level idea is similar. In round t, we have a covariance matrix $\Lambda_t := \sum_{\tau=1}^{t-1} x_{a_\tau} x_{a_\tau}^\top + I$ and an estimate $\hat{\theta}_t = \Lambda_t^{-1} \sum_{\tau=1}^{t-1} x_{a_\tau} r_\tau$. Set $\beta := \Theta(\sqrt{d\log(T/\delta)})$. Then we eliminate actions according to

$$\mathcal{A}_t := \left\{ a \in \mathcal{A} : \forall b \in \mathcal{A}, \left\langle \hat{\theta}_t, x_b - x_a \right\rangle \leq \beta \cdot \|x_a - x_b\|_{\Lambda_t^{-1}} \right\} \tag{3}$$

For action selection, we find a distribution p supported on $\mathcal{A}_i$ such that

$$\forall a \in \mathcal{A}_t : \|x_a - \mathbb{E}_{b\sim p}[x_b]\|_{\Lambda_t^{-1}}^2 \leq \mathrm{tr}(\Lambda_t^{-1} \cdot \mathbb{E}_{b\sim p}[x_b x_b^\top]). \tag{4}$$

Call that distribution p_t. We sample $a_t \sim p_t$, play it and proceed to the next round. Note that the problem in Eq. (4) is always feasible, although we will not prove this here.

For this question, you may assume that $\|\hat{\theta}_t - \theta^\star\|_{\Lambda_t} \leq \beta$ for all $t \in [T]$. More generally, you need not prove any concentration statement that you plan to use, since I do not want to you spend too much time on this aspect of the analysis. Just state what random variable you hope is concentrating to what quantity.

   (a) First, show that, by our choice of β, Eq. (3) never eliminates the optimal action $\mathrm{argmax}_a \langle x_a, \theta^\star\rangle$.

   (b) Use Eq. 3 and (4) to bound the regret on round t, namely $\max_{a\in\mathcal{A}} \langle \theta^\star, x_a - \mathbb{E}_{b\sim p_t}[x_b]\rangle$, in terms of β, Λ_t^{-1} and $M_t = \mathbb{E}_{b\sim p_t}[x_b x_b^\top]$.

   (c) Show that the algorithm's regret is $\tilde{O}(d\sqrt{T})$. You may use, without proof, the following inequality

$$\sum_t \mathrm{tr}(\Lambda_t^{-1} M_t) \leq 2 \sum_t \mathrm{tr}(\Lambda_t^{-1} x_{a_t} x_{a_t}^\top) + 8\log(2/\delta).$$

4. **Adapting to Misspecification with SquareCB.** SquareCB and most of the algorithms we have seen rely on a notion of realizability, where the expected rewards can be predicted by some function $f^\star$ in your function class $\mathcal{F}$. This assumption is unlikely to be satisfied in practice, so minimally it would be nice if our algorithms were robust to small violations of this assumption. We will study this in this question.

Suppose we are in a contextual bandits setting with function class $\mathcal{F}$, but where the mean rewards $\mu : \mathcal{X} \times \mathcal{A} \to [0,1]$ does not belong to $\mathcal{F}$. Instead assume that there is a *known* ε such that

$$\min_{f^\star \in \mathcal{F}} \sup_{x,a} |f^\star(x,a) - \mu(x,a)| \leq \varepsilon,$$

that is there is some function $f^\star \in \mathcal{F}$ that is pointwise close to the mean-reward μ. In this problem, we will study a small variant of SquareCB under misspecification.

   (a) First show that, when the oracle is given sequence $\{(x_t, a_t, r_t(x_t, a_t))\}_{t=1}^T$ where $\mathbb{E}[r_t(x_t, a_t) \mid x_t, a_t] = \mu(x,a)$, we have the bound

$$\mathbb{E}\left[\sum_{t=1}^T (\hat{y}_t(x_t, a_t) - f^\star(x_t, a_t))^2\right] \leq 2\mathbb{E}\left[\mathrm{Reg}_{\mathrm{sq}}(T)\right] + 4\varepsilon^2 T$$

   where $\hat{y}_t(x_t, a_t)$ are the regression oracle's predictions.

(b) Using the above fact, show that by changing simply by changing the learning rate γ in SquareCB, it enjoys a regret bound of

$$\mathrm{Regret}_T \leq O(\sqrt{AT\mathrm{Reg}_{\mathrm{sq}}(T)} + \varepsilon T\sqrt{A}).$$