CS7545, Fall 2018: Machine Learning Theory - Homework #4

Jacob Abernethy, Bhuvesh Kumar, and Jun-Kun Wang Due: Friday, November 16, 2018 at 4:00 pm

Homework Policy: Working in groups is fine, but *every student* must submit their own writeup. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

1) **Stochastic Bandit.** Let's consider a stochastic setting: there are two arms 1 and 2, and on each round t, arm i pays out $X_i^t \in [0,1]$, which is sampled IID from distribution D_i. The expected reward is $\mu_1 \in [0,1]$ and $\mu_2 \in [0,1]$ for D_1, D_2, respectively. Now assume that the time horizon T is known beforehand, and assume $T > 2n$. Consider the following algorithm:

Algorithm 1: Two armed bandit – explore then exploit

---

for $t = 1, \ldots, 2n$ **do**

  Pull arm $i_t = \begin{cases} 1 & 1 \le t \le n \\ 2 & n+1 \le t \le 2n \end{cases}$

  Observe $X_{i_t}^t$

end

Set $\hat{\mu}_1 = \frac{1}{n}\sum_{t=1}^{n} X_1^t$

Set $\hat{\mu}_2 = \frac{1}{n}\sum_{t=n+1}^{2n} X_2^t$

for $t = 2n+1, \ldots, T$ **do**

  Pull arm $i_t = \arg\max_{i \in \{1,2\}} \hat{\mu}_i$

end

Define the expected regret as $E[R_T] := E[\mu^* T - \sum_{t=1}^{T} \mu_{i_t}]$, where $\mu^* = \max(\mu_1, \mu_2)$ and i_t is the arm that the algorithm pulls on round t. Determine n to minimize the regret of this algorithm. What is the best choice of n? What is the bound of the regret given this number n? Important: Your regret bound can only depend on the horizon T and the gap $|\mu_1 - \mu_2|$!

2) **More Stochastic Bandit.** Consider the same 2-armed bandit setting as the previous problem. Now let us consider Algorithm 2.

Prove a regret bound for this algorithm. Important: Your regret bound can only depend on the horizon T and the gap $|\mu_1 - \mu_2|$!

(Hint: Define the event $\texttt{good}$ as being true when $|\hat{\mu}_i^t - \mu_i| \le \sqrt{\frac{2\log T}{N_i^t}}$ holds. $E[R_T] = E[R_T|\texttt{good}] \cdot Pr[\texttt{good}] + E[R_T|\neg\texttt{good}] \cdot Pr[\neg\texttt{good}]$. Assume $\texttt{good}$ is true, and let s be the last round for which $[\mathrm{LCB}_1^s, \mathrm{UCB}_1^s] \cap [\mathrm{LCB}_2^s, \mathrm{UCB}_2^s] \neq \emptyset$. What is the cumulative regret of the algorithm up to round s? What is the regret after round s?

Algorithm 2: Two armed bandit – interval intersection

for $t = 1, \ldots, n$ **do**
  for $i = 1, 2$ **do**
    $N_i^t \leftarrow \sum_{s=1}^t \mathbb{I}[i_s = i]$
    $\hat{\mu}_i^t \leftarrow \frac{1}{N_i^t} \sum_{s=1}^t \mathbb{I}[i_s = i] X_i^s$
    $\mathrm{UCB}_i^t \leftarrow \hat{\mu}_i^t + \sqrt{\frac{2 \log T}{N_i^t}}$
    $\mathrm{LCB}_i^t \leftarrow \hat{\mu}_i^t - \sqrt{\frac{2 \log T}{N_i^t}}$
  end
  if $[LCB_1^t, UCB_1^t] \cap [LCB_2^t, UCB_2^t] \neq \emptyset$ **then**
    Play arm $i_t = \begin{cases} 1 & t \text{ odd} \\ 2 & t \text{ even} \end{cases}$
  else
    Play arm $i_t = \arg\max_{i \in \{1,2\}} \hat{\mu}_i^t$
  end
end

3) **Follow-the-regularized-Leader (FTRL).** Imagine we are in the setting of Online Convex Optimization, where we have a convex decision set Ω. Recall the update of **Follow the Regularized Leader** (FTRL) is

$$w_{t+1} = \arg\min_{w \in \Omega} \eta \sum_{s=1}^{t} \ell_t(w) + R(w),$$

where $\eta > 0$ is a parameter, $\ell_t(\cdot)$ is a convex loss function on round t, and $R(\cdot)$ a strongly convex function on Ω.

Recall the definition of Bregman divergence, $D_\psi(w, v) := \psi(w) - \psi(v) - \langle \nabla \psi(v), w - v \rangle$ with respect to the function $\psi(\cdot)$. We also have an algorithm known as **Online Mirror Descent** (OMD) which is

$$w_{t+1} = \arg\min_{w \in \Omega} \eta \ell_t(w) + D_{\phi_{t-1}}(w, w_{t-1}),$$

where $\phi_1, \phi_2, \ldots$ is some sequence of convex functions.

A) Assume that we are in the unconstrained setting, $\Omega = \mathbb{R}^d$ and the sequence of functions ϕ_t is chosen as $\phi_0(\cdot) = R(\cdot)$ and $\phi_t(\cdot) = \phi_{t-1}(\cdot) + \eta \ell_t(\cdot)$. Show that OMD for this sequence is equivalent to FTRL, in that they generate the same sequence of updates.

B) Assume that we are in the unconstrained setting, $\Omega = \mathbb{R}^d$, the FTRL regularizer is $R(w) = \frac{1}{2}\|w\|^2$, and that the sequence of loss functions are linear, $\ell_t(w) = \langle \theta_t, w \rangle$. Show that, FTRL is equivalent to Online Gradient Descent:

$$w_{t+1} = w_t - \eta \nabla_{w_t} \ell_t.$$

That is, assuming they are initialized the same, show these two algorithms generate the same sequence of updates.

3) **VC dimension: Union of intervals.** What is the VC-dimension of subsets of the real line formed by the union of k intervals?

4) **VC dimension: Ellipsoids.** An n-dimensional *ellipsoid* E is defined by a center point $\mathbf{x}_0 \in \mathbb{R}^n$ and a symmetric positive semidefinite matrix $M \in \mathbb{R}^{n \times n}$, such that $\mathbf{x} \in E \iff (\mathbf{x} - \mathbf{x}_0)^\top M (\mathbf{x} - \mathbf{x}_0) \leq 1$.

(a) Give a quadratic upper bound on the VC dimension of concept class of ellipsoids in $\mathbb{R}^n$. That is, show that the VCD $= O(n^2)$.

(b) (**(hard) Challenge**) Give an exact characterization of the VC dimension of ellipsoids.