Machine Learning Theory (CS 6783)

Lecture 23-24: Decision Estimation Coefficient for Learning With Partial Information

1 Decision Making With Structured Observations (DMSO)

In the paper "The Statistical Complexity of Interactive Decision Making" by Foster et al, the general problem of DMSO was introduced that is rich enough to capture not just multi-armed bandit, linear bandit and contextual bandit problems but also more complex problems like Reinforcement Learning (RL). Let us first introduce the learning problem. We get to interact with nature n times as

For $t = 1$ to n

Learner picks policy $\pi_t \in \Pi$

Nature provides reward $r_t \in \mathcal{R}$ for that instance for the chosen policy π_t and provides observations $o_t \in \mathcal{O}$ (a general observation space).

Learner receives reward r_t and observes both reward r_t and observation o_t.

For this lecture as in the paper we focus on the case that rewards and observations are produced on every round using a fixed stochastic model $M^* : \Pi \mapsto \Delta(\mathcal{R} \times \mathcal{O})$. That is, given policy π_t, $(r_t, o_t) \sim M^*(\pi_t)$. Of course M^* is unknown to the learner. For a round t with reward $r_t \in \mathcal{R}$, we assume that the total reward for that round is give by $R_t = \text{TotalRew}(r_t)$ where TotalRew is a fixed function. In most cases, r_t is typically just a number which is the total reward for that round itself, and in the case of episodic RL, r_t is a vector of rewards and total reward for the round is simply the sum of the coordinates for the vector.

Our goal is to minimize expected regret (expectation over draws from M^*) given by

$$\text{Reg}_n = \sup_{\pi \in \Pi} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{(r_t^*, o_t^*) \sim M^*(\pi)} \left[\text{TotalRew}(r_t^*)\right] - \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{(r_t, o_t) \sim M^*(\pi_t)} \left[\text{TotalRew}(r_t)\right]$$

To make our lives easier, let us introduce the notation for expected total reward under a given stochastic model M and policy π as: $f^M(\pi) = \mathbb{E}_{(r_t, o_t) \sim M(\pi)} \left[\text{TotalRew}(r_t)\right]$. Also, given a model M, let $\pi^M = \operatorname{argmax}_{\pi \in \Pi} f^M(\pi)$ and let $\pi^* = \operatorname{argmax}_{\pi \in \Pi} f^{M^*}(\pi)$ and we also use the shorthand $f^* = f^{M^*}$. Notice that regret in this case is given by:

$$\text{Reg}_n = \frac{1}{n} \sum_{t=1}^{n} \left(f^*(\pi^*) - f^*(\pi_t)\right)$$

Since we shall allow for randomized strategies and we want regret to be small in expectation over our randomization, if on round t, π_t is drawn from p_t, we want to ensure that $\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[\left(f^*(\pi^*) - f^*(\pi_t)\right)\right]$ is small

An example of a problem captured by this setting in the episodic RL problem where each round consists of an episode of length H. Observation consists of the H states encountered in an episode belonging to state space $\mathcal{S}$, a policy $\pi : \mathcal{S} \mapsto [N]$ is a mapping that picks one of N actions $\pi(s)$ for each state $s \in \mathcal{S}$. The observation set is simply $\mathcal{O} = \mathcal{S}^H$ and the reward $r_t = (r_t[1], \ldots, r_t[H])$ the H rewards over the episode t we obtain. Model M^* specifies the initial state distribution for the first step in the episode and the Markov Decision Process given by transition kernel $T : \mathcal{S} \times [N] \mapsto \Delta(\mathcal{S})$ that specifies for each state given the action taken from that state, the probability of transiting to the next state. That is, given we are at state s_h on a given step h and we took action $a_h \in [N]$, $T(s, a)$ gives the distribution over $\mathcal{S}$ of which states we might transit to next. So $s_{h+1} \sim T(s_h, a_h)$. M^* also specifies distribution over rewards for taking a given action on a given state.

The main assumption made in the paper is that there is a class of models $\mathcal{M}$ that contains M^* the truth. Formally it is stated as follows.

Assumption 1 (Realizability)**.** *We assume that the stochastic model M^* belongs to a a class of models $\mathcal{M}$ known a priori to the learner.*

Definition 1 (Decision-Estimation Coefficient (DEC))**.** *The DEC at scale $\gamma > 0$ given a nominal model $\bar{M}$, is defined by*

$$\mathrm{Dec}_\gamma(\mathcal{M}, \bar{M}) = \inf_{p \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \mathbb{E}_{\pi \sim p} \left[f^M(\pi^M) - f^M(\pi) - \gamma D^2_H(M(\pi) | \bar{M}(\pi)) \right]$$

where $D^2_H(P|Q)$ is the Hellinger distance between two probability measures P and Q (see below for definition). We further define $\mathrm{Dec}_\gamma(\mathcal{M}) = \sup_{M \in \mathcal{M}} \mathrm{Dec}_\gamma(\mathcal{M}, M)$

Given two probability measures P and Q on same probability space, the Hellinger distance is given by $D^2_H(P|Q) = \int \left(\sqrt{dP} - \sqrt{dQ} \right)^2$. If for instance the probabilities are over countable alphabets the measure is $D^2_H(P|Q) = \sum_i \left(\sqrt{P_i} - \sqrt{Q_i} \right)^2$ and in the continuous case if p and q are the corresponding densities, then $D^2_H(P|Q) = \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx$. It is useful to note that $D^2_H(P|Q) \leq \mathrm{KL}(P|Q)$.

We will see that DEC of class $\mathcal{M}$ is a crucial complexity measure that yields both upper and lower bounds on regret for general DMSO problems. Before we provide these, a slight digression into online log loss regression will be useful.

2 Regret Minimization for Log-Loss

Say we have a general set $\mathcal{X}$ from which we receive our observations. The log loss game is as follows. On every round t, the learner first provide a distribution $q_t \in \Delta(\mathcal{X})$, then the observation for that round $x_t \in \mathcal{X}$ is given to us. At this point, we observe x_t and suffer the log-loss $\log(1/q_t(x_t))$. Given a class $\mathcal{F} \subseteq \Delta(\mathcal{X})$, our goal is to minimize regret w.r.t. to this class $\mathcal{F}$ given by

$$\mathrm{Reg}_n^{\mathrm{logloss}}(\mathcal{F}) = \frac{1}{n} \sum_{t=1}^n \log(1/q_t(x_t)) - \inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{t=1}^n \log(1/f(x_t))$$

That is we want an algorithm where $\ell(q_t, x_t) = \log(1/q_t(x_t))$. A generic algorithm we looked at for the case when the set of models is finite is the exponential weights algorithm where at time t, the

probability $p_t \in \Delta(\mathcal{F})$ we pick is given by $p_t(f) \propto \exp(-\eta \sum_{j=1}^{t-1} \ell(f, x_t))$. However note that if we pick f according to law p_t and since f itself is a distribution over $\mathcal{X}$, we can instead directly pick q_t as the implied distribution over $\mathcal{X}$. That is, we can pick $q_t \in \Delta(\mathcal{X})$ as $\mathbb{E}_{f \sim p_t}[f]$.

It can be shown that this algorithm works for log loss and that too with a fast rate for regret bound. In fact, we will show that the algorithm works with $\eta = 1$.

Lemma 2. *IF we use $p_t(f) \propto \exp(-\sum_{j=1}^{t-1} \ell(f, x_t))$ and set $q_t = \mathbb{E}_{f \sim p_t}[f]$ as our probability over X on round t, then for this algorithm we have that for this algorithm,*

$$\mathrm{Reg}_n(\mathcal{F}) \leq \frac{\log(|\mathcal{F}|)}{n}$$

Proof. As before, we first use soft-max to see that

$$-\inf_{f \in \mathcal{F}} \sum_{t=1}^{n} \log(1/f(x_t)) \leq \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n} \ell(f, x_t)\right)\right)$$

Now note that if we consider round n, then we have that on the last round $p_n(f) \propto \exp(-\sum_{j=1}^{n-1} \ell(f, x_j))$

$$
\begin{aligned}
&\sum_{t=1}^{n} \ell(q_t, x_t) - \inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{t=1}^{n} \log(1/f(x_t)) \\
&\qquad \leq \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right) \cdot \exp(-\ell(f, x_n))\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\frac{\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right) \cdot \exp(-\ell(f, x_n))}{\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)}\right) \\
&\qquad\qquad + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\mathbb{E}_{f \sim p_n}\left[\exp(-\ell(f, x_n))\right]\right) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\mathbb{E}_{f \sim p_n}\left[\exp(-\log(1/f(x_n)))\right]\right) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(\mathbb{E}_{f \sim p_n}\left[f(x_n)\right]\right) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) + \log\left(q_n(x_n)\right) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) - \log\left(1/q_n(x_n)\right) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n} \ell(q_t, x_t) - \ell(q_n, x_n) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right) \\
&\qquad = \sum_{t=1}^{n-1} \ell(q_t, x_t) + \log\left(\sum_{f \in \mathcal{F}} \exp\left(-\sum_{t=1}^{n-1} \ell(f, x_t)\right)\right)
\end{aligned}
$$

Repeating the same steps down from $n-1$ to 0 we get

$$\sum_{t=1}^{n} \ell(q_t, x_t) - \inf_{f \in \mathcal{F}} \frac{1}{n} \sum_{t=1}^{n} \log(1/f(x_t)) \leq \log\left(|\mathcal{F}|\right)$$

Thus we conclude the lemma. □

3 Upper Bound in terms of DEC

3.1 The Algorithm

The algorithm uses as black-box an online log-loss regression algorithm lets call this oracle algorithm Alg that takes past observations and produces a distribution over this space for the next round. Further, in view of what we saw in the previous section, we can use exponential weights algorithm for this blackbox and when used over set of models $\mathcal{M}$, we get the regret bound of $\frac{\log(|\mathcal{M}|)}{n}$. Now given access to this algorithm, Alg, our algorithm for the DMSO problem is given below:

For $t = 1$ to n **do**

1. Get from online regression oracle the model $\widehat{M}_t = \text{Alg}((\pi_1, o_1, r_1), \ldots, (\pi_{t-1}, o_{t-1}, r_{t-1}))$
2. $p_t = \underset{p \in \Delta(\Pi)}{\text{argmin}} \sup_{M \in \mathcal{M}} \mathbb{E}_{\pi \sim p} \left[f^M(\pi^M) - f^M(\pi) - \gamma D^2_H(M(\pi), \widehat{M}_t(\pi)) \right]$
3. Draw $\pi_t \sim p_t$ and receive rewards and observations $(r_t, o_t) \sim M^*(\pi_t)$

End For

3.2 The Upper Bound Sketch

Theorem 3. *Given any DMSO problem, the algorithm described above enjoys the regret bound*

$$\mathbb{E}\left[\text{Reg}_n\right] \leq \inf_{\gamma} \left\{ \text{Dec}_\gamma(\mathcal{M}) + \gamma \text{LoglossBnd}_n \right\}$$

where LoglossBnd_n *is the bound of the log loss regression blackbox algorithm. If specifically we use exponential weights and* $\mathcal{M}$ *were finite then we get*

$$\mathbb{E}\left[\text{Reg}_n\right] \leq \inf_{\gamma} \left\{ \text{Dec}_\gamma(\mathcal{M}) + \gamma \frac{\log(|\mathcal{M}|)}{n} \right\}$$

Proof Sketch.

$$\begin{aligned}
\mathbb{E}\left[\text{Reg}_n\right] &= \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[f^*(\pi^*) - f^*(\pi_t) \right] \right] \\
&= \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[f^*(\pi^*) - f^*(\pi_t) - \gamma D^2_H(M^*(\pi_t) | \widehat{M}_t(\pi_t)) \right] + \gamma \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[D^2_H(M^*(\pi_t) | \widehat{M}_t(\pi_t)) \right] \right] \\
&\leq \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} \sup_{M \in \mathcal{M}} \mathbb{E}_{\pi_t \sim p_t} \left[f^M(\pi^M) - f^M(\pi_t) - \gamma D^2_H(M(\pi_t) | \widehat{M}_t(\pi_t)) \right] \right] \\
&\qquad + \gamma \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[D^2_H(M^*(\pi_t) | \widehat{M}_t(\pi_t)) \right] \right]
\end{aligned}$$

By definition of p_t,

$$\begin{aligned}
&= \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} \inf_{p_t \in \Delta(\Pi)} \sup_{M \in \mathcal{M}} \mathbb{E}_{\pi_t \sim p_t} \left[f^M(\pi^M) - f^M(\pi_t) - \gamma D^2_H(M(\pi_t) | \widehat{M}_t(\pi_t)) \right] \right] \\
&\qquad + \gamma \frac{1}{n} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}_{\pi_t \sim p_t} \left[D^2_H(M^*(\pi_t) | \widehat{M}_t(\pi_t)) \right] \right]
\end{aligned}$$

But by definition of DEC

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\mathrm{Dec}_\gamma(\mathcal{M},\widehat{M}_t)\right] + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\pi_t\sim p_t}\left[D_H^2(M^*(\pi_t)|\widehat{M}_t(\pi_t))\right]\right] \\
&\le \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\pi_t\sim p_t}\left[D_H^2(M^*(\pi_t)|\widehat{M}_t(\pi_t))\right]\right] \\
&\le \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\pi_t\sim p_t}\left[\mathrm{KL}(M^*(\pi_t)|\widehat{M}_t(\pi_t))\right]\right] \\
&= \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\pi_t\sim p_t}\left[\mathbb{E}_{(r_t,o_t)\sim M^*(\pi_t)}\left[\log\left(\frac{M^*(\pi_t)(r_t,o_t)}{\widehat{M}_t(\pi_t)(r_t,o_t)}\right)\right]\right]\right] \\
&= \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{M^*(\pi_t)(r_t,o_t)}{\widehat{M}_t(\pi_t)(r_t,o_t)}\right)\right] \\
&= \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{1}{\widehat{M}_t(\pi_t)(r_t,o_t)}\right) - \frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{1}{M^*(\pi_t)(r_t,o_t)}\right)\right] \\
&\le \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{1}{\widehat{M}_t(\pi_t)(r_t,o_t)}\right) - \inf_{M\in\mathcal{M}}\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{1}{M(\pi_t)(r_t,o_t)}\right)\right] \\
&\le \mathrm{Dec}_\gamma(\mathcal{M}) + \gamma\mathrm{LoglossBnd}_n
\end{aligned}
$$

Setting γ to be the minimizer of the above we conclude the Theorem.

□

4 Lower Bound in terms of DEC

Theorem 4. *For any algorithm used for DMSO problems, we have the following lower bound in probability*

$$\mathrm{Reg}_n \ge \max_{\gamma>0}\min\left\{\mathrm{Dec}_{\gamma,O(\gamma/(n\log(n)))}(\mathcal{M}),\frac{\gamma}{n}\right\}$$

In the above

$$\mathrm{Dec}_{\gamma,\epsilon}) = \sup_{M\in\mathcal{M}}\mathrm{Dec}_\gamma(\mathcal{M}_\epsilon(M),M)$$

where $\mathcal{M}_\epsilon(M) = \{M'\in\mathcal{M} : f^M(\pi^M) \ge f^{M'}(\pi^{M'}) - \epsilon\}$