

# COMS6998-11: Homework 1 Solutions

Akshay Krishnamurthy  
akshay@cs.umass.edu

- Let us start by setting the RHS of Bernstein's inequality to be at most  $\delta$  and re-arranging:

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

This is a quadratic inequality in  $t$  and we can find the roots using the quadratic formula. That is somewhat tedious, so a shortcut is to decompose  $t = t_1 + t_2$  where  $t_1$  and  $t_2$  are chosen to ensure different parts of this inequality hold. Specifically we set

$$t_1 = \sqrt{2\left(\sum_{i=1}^n \mathbb{E}[X_i^2]\right) \log(1/\delta)}, \quad \text{and} \quad t_2 = \frac{2M}{3} \log(1/\delta)$$

Now, we can check

$$t^2 = t_1^2 + 2t_1t_2 + t_2^2 \geq t_1^2 + t_2t,$$

which, using the definitions of  $t_1$  and  $t_2$  is precisely the inequality we want to satisfy, after dividing by  $n$  to account for the difference between  $\bar{X}$  and  $\sum_{i=1}^n X_i$ .

For the generalization bound, fix a function  $f \in \mathcal{F}$  and define random variables  $Z_1, \dots, Z_n$  where  $Z_i = \mathbf{1}\{f(x_i) \neq y_i\}$ . Observe that  $\hat{R}_n(f) = \frac{1}{n} \sum_{i=1}^n Z_i$  and  $R(f) = \mathbb{E}[Z_i]$ . So we will apply Bernstein's inequality on the random variables

$$\hat{R}_n(f) - R(f) = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}\{f(x_i) \neq y_i\} - R(f))$$

A critical step in the argument is that the variance term that appears in Bernstein's inequality can be upper bounded by the mean:

$$\mathbb{E}\left[(\mathbf{1}\{f(x_i) \neq y_i\} - R(f))^2\right] \leq \mathbb{E}\left[\mathbf{1}\{f(x_i) \neq y_i\}^2\right] \leq \mathbb{E}\left[\mathbf{1}\{f(x_i) \neq y_i\}\right] = R(f).$$

Then, by our variant of Bernstein's inequality and a union bound, we have that with probability at least  $1 - \delta$ :

$$\forall f \in \mathcal{F} : \left| \hat{R}_n(f) - R(f) \right| \leq \sqrt{\frac{2R(f) \log(2|\mathcal{F}|/\delta)}{n}} + \frac{2 \log(2|\mathcal{F}|/\delta)}{3}.$$

Using this with  $\hat{f}_n$  along with the fact that  $\hat{f}_n$  is the ERM yields

$$\begin{aligned} R(\hat{f}_n) - R(f^*) &\leq \hat{R}_n(\hat{f}_n) - \hat{R}_n(f^*) + \sqrt{\frac{2R(\hat{f}_n) \log(2|\mathcal{F}|/\delta)}{n}} + \sqrt{\frac{2R(f^*) \log(2|\mathcal{F}|/\delta)}{n}} + \frac{4 \log(2|\mathcal{F}|/\delta)}{3} \\ &\leq \sqrt{\frac{2R(\hat{f}_n) \log(2|\mathcal{F}|/\delta)}{n}} + \sqrt{\frac{2R(f^*) \log(2|\mathcal{F}|/\delta)}{n}} + \frac{4 \log(2|\mathcal{F}|/\delta)}{3}. \end{aligned}$$

The only problem is the first term in the above bound since it depends on  $R(\hat{f}_n)$  rather than  $R(f^*)$ . On the other hand, the bound itself is showing that  $R(\hat{f}_n) \approx R(f^*)$  so intuitively we should be able to swap the  $R(\hat{f}_n)$  term on the RHS with  $R(f^*)$  without paying too much. Indeed

$$\begin{aligned} \sqrt{\frac{2R(\hat{f}_n) \log(2|\mathcal{F}|/\delta)}{n}} &= \sqrt{\frac{2 \left( R(\hat{f}_n) - R(f^*) + R(f^*) \right) \log(2|\mathcal{F}|/\delta)}{n}} \\ &\leq \sqrt{\frac{2 \left( R(\hat{f}_n) - R(f^*) \right) \log(2|\mathcal{F}|/\delta)}{n}} + \sqrt{\frac{2R(f^*) \log(2|\mathcal{F}|/\delta)}{n}} \\ &\leq \frac{R(\hat{f}_n) - R(f^*)}{2} + \frac{\log(2|\mathcal{F}|/\delta)}{n} + \sqrt{\frac{2R(f^*) \log(2|\mathcal{F}|/\delta)}{n}} \end{aligned}$$

Here the first inequality is that  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  while the second is that  $\sqrt{ab} \leq a/2 + b/2$ . Plugging this bound in above, we have

$$R(\hat{f}_n) - R(f^*) \leq \frac{R(\hat{f}_n) - R(f^*)}{2} + 2\sqrt{\frac{2R(f^*) \log(2|\mathcal{F}|/\delta)}{n}} + \frac{7 \log(2|\mathcal{F}|/\delta)}{3n}$$

By bringing the first term from the RHS to the LHS and then multiplying by 2, we obtain the result.

2. The first observation is that for any epoch  $i$  if an arm survives into epoch  $i$  then at the end of the epoch it has been played  $N_i = \sum_{j=1}^i 2^j \geq 2^i$ . In particular this means that all surviving arms have been played the same number of times.

Let  $a^* = \operatorname{argmax}_a \mu(a)$  denote the best arm. Next, we show that  $a^*$  is never eliminated. To do this, recall that we set the confidence interval such that

$$\forall t \in [T], a \in \mathcal{A} : |\hat{\mu}_t(a) - \mu(a)| \leq \operatorname{conf}_t(a)$$

We show that for any epoch  $i$ ,  $a^*$  is not eliminated. This follows since for any  $a \neq a^*$

$$\hat{\mu}(a^*) + \operatorname{conf}(a) \geq \mu(a^*) \geq \mu(a) \geq \hat{\mu}(a) - \operatorname{conf}(a),$$

where the first and third inequality are by the property of the confidence interval and the second inequality is by the optimality of  $a^*$ .

Next we claim that any arm that survives an elimination step cannot be too suboptimal. Indeed

$$\mu(a) \geq \hat{\mu}(a) - \operatorname{conf}(a) = \hat{\mu}(a) + \operatorname{conf}(a) - 2\operatorname{conf}(a) \geq \hat{\mu}(a^*) - \operatorname{conf}(a^*) - 2\operatorname{conf}(a) \geq \mu(a^*) - 2\operatorname{conf}(a^*) - 2\operatorname{conf}(a)$$

Since the confidence interval for all surviving arms is the same, we see that if arm  $a$  survives through epoch  $i$  then its suboptimality is roughly  $\sqrt{1/2^i}$ .

For the regret analysis, the bookkeeping has to be done somewhat carefully. We decompose the regret across epochs. In each epoch the regret we incur depends on the number of active arms  $|\mathcal{A}_i|$ , the epoch length, and the confidence interval. Formally the regret in epoch  $i$  is bounded by  $4|\mathcal{A}_i| \cdot 2^i \cdot \operatorname{conf}_{i-1}$  (below we will not track the constants precisely). Let  $I$  denote the index of the last *complete* epoch and let  $I_a$  denote the index of the epoch when  $a$  is eliminated (or  $I$  if  $a$  is never eliminated) and let  $N_a$  denote the number of plays for arm  $a$  up to and including epoch  $I$ .

First we claim that we do not incur too much regret in the final epoch  $I+1$ , which does not complete. Suppose that  $|\mathcal{A}_{I+1}| = k > 0$ . Then observe that  $I \leq \lfloor \log_2(T/K) \rfloor$ , since epoch  $I$  lasts for at least  $k2^I$  rounds. The regret in epoch  $I+1$  can be crudely upper bounded by

$$k \cdot 2^{I+1} \cdot \operatorname{conf}_I \leq ck2^{I+1} \cdot \sqrt{\frac{\log(AT/\delta)}{2^I}} \leq ck\sqrt{2^I \log(AT/\delta)} \leq c\sqrt{kT \log(AT/\delta)} \leq O(\sqrt{AT \log(AT/\delta)})$$

For the other epochs we have

$$\begin{aligned}
\sum_{i=1}^I |\mathcal{A}_i| \cdot 2^i \cdot \text{conf}_{i-1} &\leq c \sum_{i=1}^I |\mathcal{A}_i| \sqrt{2^i \cdot \log(AT/\delta)} \\
&\leq c \sum_a \sum_{i=1}^{I_a} \sqrt{2^i \cdot \log(AT/\delta)} \\
&\leq c \sum_a \sqrt{N_a \cdot \log(AT/\delta)} \\
&\leq c \sqrt{AT \log(AT/\delta)}.
\end{aligned}$$

To pass to the second line, we simply re-arrange the sum so that for each arm  $a$  we consider all the epochs until  $a$  is eliminated. Then we bound this geometric series (the constant  $c$  changes from line to line). Finally we use Cauchy-Schwarz since  $\sum_a N_a \leq T$ .

3. For item (a) our choice of  $\beta$  ensures that  $\|\hat{\theta}_t - \theta^*\|_{\Lambda_t} \leq \beta$ , just as we saw in class. Now considering  $a^*$  and some other action  $b$ , we have

$$\begin{aligned}
\langle \hat{\theta}_t, x_b - x_{a^*} \rangle &= \langle \hat{\theta}_t - \theta^*, x_b - x_{a^*} \rangle + \langle \theta^*, x_b - x_{a^*} \rangle \\
&\leq \langle \hat{\theta}_t - \theta^*, x_b - x_{a^*} \rangle \\
&\leq \|\hat{\theta}_t - \theta^*\|_{\Lambda_t} \cdot \|x_b - x_{a^*}\|_{\Lambda_t^{-1}} \\
&\leq \beta \|x_b - x_{a^*}\|_{\Lambda_t^{-1}}
\end{aligned}$$

Here the first inequality uses the optimality of  $a^*$  (for the true parameter  $\theta^*$ ). The second is Cauchy-Schwarz and the third uses our guarantee on  $\hat{\theta}_t$ . This proves that  $a^*$  is never eliminated.

For item (b), the total regret across all rounds is

$$\sum_{t=1}^T \langle \theta^*, x_{a^*} - x_{a_t} \rangle \leq \sum_{t=1}^T \sum_a p_t(a) \langle \theta^*, x_{a^*} - x_a \rangle + O(\sqrt{T \log(1/\delta)}),$$

where the latter holds with probability  $1 - \delta$  by an application of Azuma's inequality. Now let us focus on one of the terms in the first sum:

$$\begin{aligned}
\sum_a p_t(a) \langle \theta^*, x_{a^*} - x_a \rangle &= \langle \theta^*, x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a] \rangle \\
&= \langle \theta^* - \hat{\theta}_t, x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a] \rangle + \langle \hat{\theta}_t, x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a] \rangle \\
&\leq \beta \cdot \|x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a]\|_{\Lambda_t^{-1}} + \langle \hat{\theta}_t, x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a] \rangle
\end{aligned}$$

The first term is at a point where we can apply Eq. (4), so let us continue with the third term. By applying Eq. (3) and then the triangle inequality we obtain

$$\langle \hat{\theta}_t, x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a] \rangle \leq \beta \sum_a p_t(a) \|x_{a^*} - x_a\|_{\Lambda_t^{-1}} \leq \beta \|x_{a^*} - \mathbb{E}_{a \sim p_t}[x_a]\|_{\Lambda_t^{-1}} + \beta \sum_a p_t(a) \|\mathbb{E}_{b \sim p_t}[x_b] - x_a\|_{\Lambda_t^{-1}}$$

All of these terms are bounded by Eq. (4). Together, we obtain the bound

$$\sum_a p_t(a) \langle \theta^*, x_{a^*} - x_a \rangle \leq 3\beta \sqrt{\text{tr}(\Lambda_t^{-1} \cdot M_t)}$$

For item (c) after the application of Azuma's inequality from above, and using our answer to part (b) the regret is bounded as

$$\sum_{t=1}^T \langle \theta^*, x_{a_t} - x_{a_t^*} \rangle \leq 3\beta \sum_{t=1}^T \sqrt{\text{tr}(\Lambda_t^{-1} \cdot M_t)} + O(\sqrt{T \log(1/\delta)})$$

We will focus on the first term. By Cauchy-Schwarz and then by using the concentration inequality provided, we bound this term as

$$\begin{aligned} \sum_{t=1}^T \sqrt{\text{tr}(\Lambda_t^{-1} \cdot M_t)} &\leq \sqrt{T} \cdot \sqrt{\sum_{t=1}^T \text{tr}(\Lambda_t^{-1} \cdot M_t)} \\ &\leq \sqrt{T} \cdot \sqrt{\sum_{t=1}^T \text{tr}(\Lambda_t^{-1} x_{a_t} x_{a_t}^\top) + 8 \log(2/\delta)} \\ &\leq \sqrt{T} \cdot \sqrt{\sum_{t=1}^T x_{a_t}^\top \Lambda_t^{-1} x_{a_t} + O(\sqrt{T \log(1/\delta)})}. \end{aligned}$$

Now we can apply the elliptical potential lemma, since  $\Lambda_{t+1} = \Lambda_t + x_{a_t} x_{a_t}^\top$ . The only minor difference is that we don't have the  $\min(1, \cdot)$  in our sum. However since we started with  $\Lambda_1 = I$ , this does not affect the argument too much. In particular the only change is in the first step, where we use the Woodbury identity, Here we get

$$x_{a_t}^\top \Lambda_{t+1}^{-1} x_{a_t} = \frac{x_{a_t}^\top \Lambda_t^{-1} x_{a_t}}{1 + x_{a_t}^\top \Lambda_t^{-1} x_{a_t}} \geq (1 + B^2)^{-1} x_{a_t}^\top \Lambda_t^{-1} x_{a_t}$$

which follows since  $\Lambda_t^{-1} \preceq I$  and we assume that the features are bounded as  $\|x_{a_t}\|_2 \leq B$ . Using the rest of the elliptical potential as is, we can conclude that

$$\sum_{t=1}^T x_{a_t}^\top \Lambda_t^{-1} x_{a_t} \leq \frac{1}{1 + B^2} d \log(1 + TB^2/d),$$

and using this yields the desired regret bound.

4. For part (a) observe that

$$\begin{aligned} &\mathbb{E} [(\hat{y}_t(x_t, a_t) - r_t)^2 - (f^*(x_t, a_t) - r_t)^2] \\ &= \mathbb{E} [\hat{y}_t(x_t, a_t)^2 - f^*(x_t, a_t)^2 - 2r_t(\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))] \\ &= \mathbb{E} [\hat{y}_t(x_t, a_t)^2 - f^*(x_t, a_t)^2 - 2\mu_t(x_t, a_t)(\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))] \\ &= \mathbb{E} [(\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))^2 - 2(\mu_t(x_t, a_t) - f^*(x_t, a_t))(\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))] \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=1}^T (\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))^2 \right] &\leq \mathbb{E} [\text{Reg}_{\text{sq}}(T)] + \mathbb{E} \left[ \sum_{t=1}^T 2(\mu_t(x_t, a_t) - f^*(x_t, a_t))(\hat{y}_t(x_t, a_t) - f^*(x_t, a_t)) \right] \\ &\leq \mathbb{E} [\text{Reg}_{\text{sq}}(T)] + \mathbb{E} \left[ \sum_{t=1}^T 2(\mu_t(x_t, a_t) - f^*(x_t, a_t))^2 + \sum_{t=1}^T \frac{1}{2} (\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))^2 \right] \\ &\leq \mathbb{E} [\text{Reg}_{\text{sq}}(T)] + 2\varepsilon^2 T + \mathbb{E} \left[ \sum_{t=1}^T \frac{1}{2} (\hat{y}_t(x_t, a_t) - f^*(x_t, a_t))^2 \right] \end{aligned}$$

Re-arranging this proves the result.

For the part (b) we need to prove a “per-round” inequality, but first let us see what the per-round inequality should be. The first thing to do is to move from the true reward function  $\mu$  to the function  $f^*$  that is in our function class. For this, observe that for any  $x_t$  and  $a_t$

$$\begin{aligned} \max_a \mu(x_t, a) - \mu(x_t, a_t) &= \max_a \mu(x_t, a) - \max_a f^*(x_t, a) + \max_a f^*(x_t, a) - f^*(x_t, a_t) + f^*(x_t, a_t) - \mu(x_t, a_t) \\ &\leq 2\varepsilon + \max_a f^*(x_t, a) - f^*(x_t, a_t), \end{aligned}$$

where the inequality holds by our assumption on  $f^*$  and  $\mu$  (and a similar proof to what we saw in the contraction lemma to relate the two max’s).

So from now on we can work with  $\mathbb{E}_{a_t \sim p_t} [\max_a f^*(x_t, a) - f^*(x_t, a_t)]$ . Let us condense the notation and drop dependence on  $x_t$  and  $t$  altogether. Let  $a^* = \operatorname{argmax}_a f^*(a)$  and let  $y$  denote the vector of our predictions and let  $b = \operatorname{argmax}_a y(a)$ . So now we want to bound

$$\mathbb{E}_{a \sim p} [f^*(a^*) - f^*(a)] - \frac{\gamma}{4} \mathbb{E}_{a \sim p} [(y(a) - f^*(a))^2],$$

and exactly our analysis for SquareCB shows that this is  $O(A/\gamma)$ . Putting things together, we have

$$\begin{aligned} \operatorname{Regret}(T) &:= \sum_{t=1}^T \max_a \mu(x_t, a) - \mathbb{E}_{a_t \sim p_t} [\mu(x_t, a_t)] \\ &\leq 2T\varepsilon + \sum_{t=1}^T \max_a f^*(x_t, a) - \mathbb{E}_{a_t \sim p_t} f^*(x_t, a_t) \\ &\leq 2T\varepsilon + \frac{AT}{\gamma} + \frac{\gamma}{4} \sum_{t=1}^T \mathbb{E}_{a \sim p_t} [(\hat{y}_t(x_t, a) - f^*(x_t, a))^2] \\ &\lesssim 2T\varepsilon + \frac{AT}{\gamma} + \frac{\gamma}{4} (2\operatorname{Reg}_{\text{sq}}(T) + 2\varepsilon^2 T) \end{aligned}$$

Here the  $\lesssim$  notation hides a concentration argument to relate  $\operatorname{Reg}_{\text{sq}}(T)$  to its expectation. Now if we optimize for  $\gamma$  we get  $\gamma = \Theta\left(\sqrt{\frac{AT}{\operatorname{Reg}_{\text{sq}}(T) + 2\varepsilon^2 T}}\right)$  and plugging this in yields

$$O\left(T\varepsilon + \sqrt{AT(\operatorname{Reg}_{\text{sq}}(T) + 2\varepsilon^2 T)}\right) \leq O\left(\varepsilon T\sqrt{A} + \sqrt{AT\operatorname{Reg}_{\text{sq}}(T)}\right)$$