6.S978 Graphs, Linear Algebra, and Optimization – Fall 2015

Problem Set 2

Out: November 4, 2015 *Due: November 25, 2015*

(1) *Submit the solutions as a PDF file (typeset in LaTeX) by emailing it to Aleksander.*

(2) *You can solve the problems in collaboration with one other person, but your writeup has to be prepared independently. Also, please provide the name of the collaborator.*

Problem 1. Recall our toy model of stock market prediction (see Section 2 in the notes from Lecture 9). Prove that, in the worst case, no *deterministic* prediction algorithm for this setting can make less than $2M^*(T) + \lfloor \log_2 n \rfloor$ mistakes, whenever $n \geq 2$. Here, n is the size of the expert set and $M^*(T)$ is the number of mistakes made by the best expert over T rounds.

Note: This means that it is essential to employ randomness to obtain the performance guarantee of the randomized weighted majority algorithm.

Problem 2. Show that for any (non-empty) subset S of n experts, the multiplicative weights update method suffers a total loss $L(T)$ of at most

$$L(T) \leq \max_{i \in S} \left(\sum_t l_i^t + \eta \sum_t |l_i^t| \right) + \frac{\ln \frac{n}{|S|}}{\eta},$$

where $0 \leq \eta \leq \frac{1}{2}$ and each $l_i^t \in [-1, 1]$.

Note: This means that the multiplicative weights update method has (slightly) better performance when there are many – instead of only one – "good" experts.

Problem 3. Consider the learning from expert advice framework (see Section 5 in the notes from Lecture 9) with n experts. Recall that even though we allowed the multiplicative weights update method to switch between experts in each round, we compared its performance only to the best *fixed* expert.

(a) How important was it that this best expert we compare ourselves to is fixed?

To answer this question, show that, for any algorithm ALG, one can construct an instance of the learning from expert advice framework such that: (1) the total loss $L_{ALG}(T)$ of this algorithm after T rounds will be at least $T(1 - 1/n)$; (2) the total loss $\hat{L}^*(T)$ of the best "changing" expert is 0. (Here, the loss of the changing expert is simply $\hat{L}^*(T) := \sum_{t=1}^{T} \min_i l_i^t$.)

(b) Consider a situation in which we want to be able to bound our performance not only wrt the best fixed expert but also wrt the best "alternation" of *at most* k experts. That is, we want to compare ourselves to the loss

$$L_k^*(T) := \min_{1 \leq t_1 \leq \ldots t_{k-1} \leq T} \sum_{l=0}^{k-1} \min_i \sum_{t=t_l}^{t_{l+1}-1} l_i^t,$$

where we used a convention that $t_0 = 1$ and $t_k = T + 1$.

Design an algorithm for this scenario whose regret $\text{Regret}_k(T)$ is at most

$$\text{Regret}_k(T) := L(T) - \hat{L}_k^*(T) = O(\sqrt{kT \log Tn}).$$

(Assume that the number of rounds T is known in advance.)

Note: The optimal regret bound one can get here is $O(\sqrt{kT \log n})$, but you do not need to show that.

Problem 4. Let $\mathcal{E}$ be the energy of a unit electrical s-t-flow wrt resistances r. Recall that in Lecture 4 we showed that

$$\frac{1}{2} \cdot \mathcal{E} = \min_{h \in \mathbb{R}^m} \max_{\varphi \in \mathbb{R}^n} \mathcal{L}(h, \varphi),$$

where

$$\mathcal{L}(h, \varphi) := \frac{1}{2} h^T R h - (Bh - \chi_{st})^T \varphi$$

is the Lagrangian of the unit electrical s-t-flow problem and R is an m-by-m diagonal matrix with each diagonal entry R_{ee} equal to r_e. Also, we mentioned that strong duality holds for such Lagrangian, i.e., that we have that

$$\min_{h \in \mathbb{R}^m} \max_{\varphi \in \mathbb{R}^n} \mathcal{L}(h, \varphi) = \max_{\varphi \in \mathbb{R}^n} \min_{h \in \mathbb{R}^m} \mathcal{L}(h, \varphi).$$

(a) Argue that strong duality implies that

$$\frac{1}{2} \cdot \mathcal{E} = \max_{\varphi \in \mathbb{R}^n} \varphi^T \chi_{s,t} - \frac{1}{2} \varphi^T L \varphi,$$

where $L = BR^{-1}B^T$ is the Laplacian of the underlying graph with edge weights given by r_e^{-1}s.

(b) Use (a) to argue that

$$\frac{1}{2} \cdot \mathcal{E} = \widehat{\varphi}^T \chi_{s,t} - \frac{1}{2} \widehat{\varphi}^T L \widehat{\varphi},$$

where $\widehat{\varphi}$ are vertex potentials that induce the electrical flow $\hat{f}$ via Ohm's law, i.e., $\widehat{\varphi}$ are such that

$$\hat{f} = R^{-1} B^T \widehat{\varphi}.$$

General hint: Use optimality conditions.

Problem 5. Let $\hat{f}$ be a unit electrical s-t-flow in the graph G wrt resistances r and let $\mathcal{E} := \sum_{e \in E(G)} r_e \hat{f}_e^2$ be the energy of that flow. Also, let $\bar{e}$ be an edge in G such that its contribution $r_{\bar{e}} \hat{f}_{\bar{e}}^2$ to the energy of $\hat{f}$ constitutes at least a δ-fraction of that energy, i.e.,

$$r_{\bar{e}} \hat{f}_{\bar{e}}^2 \geq \delta \mathcal{E}.$$

Use the conclusions of Problem 4 to argue that, if $\mathcal{E}'$ is the energy of a unit electrical s-t-flow wrt resistances r_e in the graph G with the edge $\bar{e}$ removed, then

$$\mathcal{E}' \geq (1 + \delta) \mathcal{E}.$$

(Think of removing the edge $\bar{e}$ as setting its resistance $r_{\bar{e}}$ to be infinite.)