Minimalizacja niezadowolenia, wieloramienny bandyta
Algorytmiczna Teoria Gier IV 2010

Ćwiczenie 1. Przedstaw dowód twierdzenia 4.12 lub 4.13 z książki.

Ćwiczenie 2. Rozszerz sformułowanie algorytmu `PW` i twierdzenia 4.6 z książki tak, aby obejmowało maksymalizację dowolnych wypłat ograniczonych od dołu i pokrótce omów niewielkie potrzebne zmiany w dowodzie. Uniezależnij wynik od znajomości górnego ograniczenia (sum kwadratów) wypłat – przedstaw modyfikację algorytmu (zapoznaj się pobieżnie z dowodem). Patrz *Improved second-order bounds for prediction with expert advice*, lemat 2 i paragraf 3.1, http://arxiv.org/pdf/math.ST/0602629.

Ćwiczenie 3. W definicji straty algorytmu H: $L_H^T = \sum_{t=1}^T \sum_{i=1}^N p_i^t l_i^t$, w modelu minimalizacji strat, posługujemy się wartością oczekiwaną. Zamiast tego, zdefiniujmy zmienną losową $\hat{L}_H^T = \sum_{t=1}^T \sum_{i=1}^N l_{\hat{i_t}}^t$, gdzie $\hat{i_t} \sim \overline{p_i^t}$ (zmienna losowa z rozkładu zadanego przez algorytm H w chwili t). Znajdź nietrywialne oszacowanie straty z prawdopodobieństwem $1-\delta$: $P\{\hat{L}_H^T > f(T,\delta)\} < \delta$ dla małych dodatnich δ, dla algorytmu $H = \mathtt{PW}'$ minimalizacji niezadowolenia, z pełną informacją, nazwanego w książce `PolynomialWeights` – lub jego modyfikacji (np. zmniejszającej wariancję strat algorytmu). (Patrz paragraf 4.3 w *Improved second-order bounds for prediction with expert advice*. Paragraf 6 pracy *The non-stochastic multi-armed bandit problem* znajduje takie oszacowanie dla algorytmu częściowej informacji `Exp3.P`.)

Ćwiczenie 4. Przedstaw dowód twierdzenia 4.17 z książki (oszacowanie dla algorytmu minimalizacji niezadowolenia z częściową informacją `MAB`).

Ćwiczenie 5. Przedstaw algorytm `Exp3` z pracy *The non-stochastic multi-armed bandit problem* (http://cseweb.ucsd.edu/~yfreund/papers/bandits.pdf) – paragraf 3. Uzasadnij jego postać porównując go z algorytmem `MAB`, przytaczając odpowiednie fragmenty dowodów.

Ćwiczenie 6. Algorytmy minimalizacji niezadowolenia przy częściowej informacji (tzn. dla problemu wielorękiego bandyty) są wrażliwe na dużą ilość akcji ze względu na konieczność oszacowania strat/wypłat dla każdej akcji osobno. Załóżmy że mamy do dyspozycji dużo ekspertów $i \in \{1, ..., N\}$ dostarczających rozkładu prawdopodobieństw $\xi^i(t)$ dla małej liczby akcji $j \in \{1, ..., K\}$. Wzorując się na algorytmie `Exp4` z *The non-stochastic multi-armed bandit problem*, zmodyfikuj algorytm `MAB` (bazującego na `PW`) oraz wyprowadź dla niego górne oszacowanie na oczekiwane niezadowolenie w klasie porównawczej, które zależy tylko logarytmicznie od liczby ekspertów.

Ćwiczenie 7. Trudnością strategii dokonującej wyborów $j_1, ..., j_T$ nazywamy liczbę zmian akcji w ciągu j_t: $1 + |\{1 \leqslant t < T : j_t \neq j_{t+1}\}|$. Wyprowadź nietrywialne oszacowanie niezadowolenia w porównaniu z dowolną strategią dla algorytmu `MAB` (bazującego na `PW`) lub jego modyfikacji, zależne od trudności porównywanej strategii. (Patrz paragraf 8 artykułu *The non-stochastic multi-armed bandit problem* – algorytm `Exp3.S`.)

Wybieranie deterministycznie najlepszego eksperta w modelu z pełną informacją przynosi niezadowolenie względem najlepszego eksperta rzędu „ilość zmian lidera”, w najgorszym przypadku $T/2$ dla strat z przedziału $[0, 1]$. Okazuje się, że strategię tę można w prosty sposób poprawić.

Ćwiczenie 8. Strategia „podążaj za zabużonym liderem” polega na przechowywaniu strat do chwili t poszczególnych ekspertów/akcji $L_{i,t}$, i wybieraniu akcji $\mathrm{argmin}_{i=1,...,N}(L_{i,t-1} + Z_{i,t})$ dla pewnych zmiennych losowych $Z_{i,t}$. Gdy $Z_{i,t}$ są niezależnymi zmiennymi z rozkładu jednostajnego na $[0, \Delta]$, wtedy oczekiwane niezadowolenie algorytmu jest ograniczone przez

$$L_T - \min_{i=1,...,N} L_{i,T} \leqslant \Delta + \frac{NT}{\Delta}$$

Skonstruuj środowisko (zależne od Δ), w którym strata tego algorytmu jest możliwie duża (tzn. znajdź nietrywialne, być może pasujące, dolne ograniczenie).

Gdy rozkład zaburzeń skupimy wokół najlepszych akcji, zmniejszymy wrażliwość strategii na ilość ekspertów – czy widać to dla skonstruowanego przez ciebie środowiska? Dla strategii „podążaj za zabużonym liderem" z rozkładem zaburzeń $\boldsymbol{Z}_t = (Z_{1,t}, \ldots, Z_{N,t})$ o gęstości $f(\boldsymbol{z}) = (\eta/2)^N e^{-\eta\|\boldsymbol{z}\|_1}$ (gdzie $\|\boldsymbol{z}\|_1 := \sum_{i=1}^N |z_i|$), ograniczenie górne zależy tylko logarytmicznie od ilości ekspertów (*Prediction, Learning and Games* rozdział 4.3).