4EK421 – zadání úkolu – opakované hry

Zadání

Uvažujme bimaticovou hru s následujícími výplatními maticemi

$$X_1 = \begin{pmatrix} 5 & 3 & 1 & 3 \\ 3 & 4 & 3 & 1 \\ 3 & 1 & -3 & 5 \\ 4 & 2 & -1 & 4 \end{pmatrix}, \text{ a } \qquad X_2 = X_1^T = \begin{pmatrix} 5 & 3 & 3 & 4 \\ 3 & 4 & 1 & 2 \\ 1 & 3 & -3 & -1 \\ 3 & 1 & 5 & 4 \end{pmatrix}$$

Řekněme, že se hra bude hrát opakovaně. Hráči se snaží maximalizovat normalizovnou současnou hodnotu výhry k okamžiku, kdy se hraje nultá hra, přičemž pro diskontování budoucích výher a normalizaci se používá určitý neznámý diskontní faktor $\delta < 1$ (přesněji viz definice ve slidech). Cílem je formulovat předpis, podle kterého má řádkový hráč volit strategii v t-tém opakování hry v závislosti na strategiích v minulých hrách, pro $t = 0, 1, \ldots$. Pro označení strategií v předpisech používejme čísla 0, 1, 2 a 3, kde 0 značí 1. řádek, 1 značí 2. řádek, 2 značí 3. řádek a 3 značí 4. řádek. V předpisech označujme symbolem a_t strategii řádkového hráče v t-té hře a symbolem b_t strategii sloupcového hráče v t-té hře.

Příklad předpisu, který v prvních 11 hrách (hry jsou indexované od 0) vybere 1. řádek (nebo sloupec) a dále střídá 2., 3. a 4. řádek:

$$a_t = \begin{cases} 0 & \text{pokud } t \leq 10, \\ 1 & \text{pokud } t > 10 \wedge t \bmod 3 = 2, \\ 2 & \text{pokud } t > 10 \wedge t \bmod 3 = 0, \\ 3 & \text{pokud } t > 10 \wedge t \bmod 3 = 1 \end{cases}$$

kde $\bmod$ je infixový operátor vracející zbytek po dělení. (Poslední tři řádky předpisu lze nahradit řádkou $1 + (t + 1 \bmod 3)$ pokud $t > 10$.

Příklad předpisu, který v první hře zahraje 2. řádek a v každé další kopíruje bezprostředně předcházející strategii protihráče:

$$a_t = \begin{cases} 1 & \text{pokud } t = 0, \\ b_{t-1} & \text{pokud } t > 0. \end{cases}$$

Další příklady předpisů (pro hru s 2×2 strategiemi) naleznete na slidech `zde`.

Odevzdání a hodnocení

Úkol se odevzdává do příslušné odevzdávárny v InSIS. Za zaslání úkolu lze získat 7 bodů. Počet bodů závisí na nejlepším z výsledků odevzdaného předpisu v porovnání předpisů níže. Za korektní předpis budou vždy uděleny alespoň 3 body. **Předpis**

však musí být zároveň slovně vysvětlen. Umístí-li se předpis alespoň v jednom z 27 dílčích pořadí:

- mezi prvními 2 předpisy, získá navíc 1 body,
- v první třetině pořadí předpisů, získá navíc 1 body,
- v první polovině pořadí předpisů, získá navíc 1 body.

Například, bude-li předpis alespoň v jednom z kritérií a pro nějakou kombinaci diskotního faktoru a časového horizontu nejlepší ze všech předpisů a bude-li korektně odevzdaný, získá jeho autor celkem $3+2+3+2 = 10$ bodů, nezávisle na tom, jakých výsledků bude dosahovat pro ostatní diskontní faktory, kritéria a časové horizonty.

Pro získání maximálního počtu bodů je tedy postačující, exceluje-li předpis v některém ze srovnávacích kritérií a pro nějaké hodnoty diskontního faktoru a počtu opakování.

Porovnání předpisů a bonusové body

Předpis v pythonu. V tomto online python notebooku lze najít příklady předpisů, které byly vytvořeny v minulých semestrech (pro jiné výplatní matice). Každý předpis je reprezentován jednou funkcí, která na vstupu dostane seznamy strategií napozorované v minulých opakováních hry a která vrátí číslo 0, 1, 2 nebo 3 v závislosti na tom, jakou strategii má předpis v aktuálním opakování vybrat. Funkce může přistupovat ke globálním proměnných `A` (výplatní matice; předpis je tvořen pro řádkového hráče) a `delta` (hodnota diskontního faktoru). Pro přístup k výplatní matici lze použít i funkci `vyplata`.

Předpis, který bude odevzdán ve formě pythonovské funkce s výše naznačeným rozhraním, obdrží 1 bod navíc. Nejasnosti prosím předem konzultujme.

Porovnání předpisů. Po odevzdání všech předpisů budou předpisy porovnány. Na začátku náhodně vygenerujeme čísla T_1, T_2, T_3 a $\delta_1, \delta_2, \delta_3$. Porovnání bude spočívat v tom, že každou dvojici předpisů necháme hrát proti sobě pro každou kombinaci $T = T_i$ a $\delta = \delta_j$ pro $i \in \{1,2,3\}$ a $j \in \{1,2,3\}$. Každá dvojice předpisů tak vůči sobě bude hrát 9krát. Jedné instanci opakované hry pro dvojici předpisů říkejme *párové srovnání*. Očekávejme, že T bude moci být číslo v intervalu $[3, 1000]$ a δ číslo v intervalu $[0.1, 1-10^{-10}]$. Hodnota T nicméně není hráčům známá a není možné ji použít v předpisech.

Pro každou kombinaci T a δ vytvoříme 3 uspořádání předpisů. Jednotlivá uspořádání budou řadit podle následujících kriteriálních hodnot (tj. příslušná hodnota bude spočtena pro každý předpis):

1. počet vyhraných párových srovnání (pro dané T a dané δ), tj. počet párových srovnání, ve kterých předpis dosáhl vyšší výplaty než předpis, se kterým se srovnává;
2. součet výplat předpisu přes všechna párová srovnání (pro dané T a dané δ), kterých se předpis účastní;
3. součet rozdílů výplat přes všechna párová srovnání (pro dané T a dané δ), kterých se předpis účastní.

Získáme 27 uspořádání předpisů. Ty blíže neurčeným způsobem agregujeme do jednoho uspořádání. Autoři předpisů, které budou na předních místech, získají kromě řádných bodů také blíže neurčený počet bonusových bodů.