

Výzkumník strojového učení pro reklamní systémy

Úlohy pro přijímací řízení

Děkujeme, že jste se přihlásili do přijímacího řízení na pozici výzkumníka v reklamních systémech. Připravili jsme pro Vás praktické úlohy, které svým charakterem odráží část náplně naší práce.

Pokyny pro vypracování:

- Pro řešení úlohy využijte Python a jeho knihovny.
- Nejde nám o vyčerpávající výčet všech možných řešení, spíše o zvolení vhodné technologie a postupu. Potřebnou časovou dotaci odhadujeme na jedno až dvě odpoledne.
- Některé úlohy mají více řešení. Zkuste zvolit takové, které Vám přijde nejvhodnější a okomentujte volbu.
- V případě nejasností kontaktujte HR specialistu, který s Vámi komunikuje.
- Výstup by měl obsahovat:
 - zip soubor obsahující py nebo ipynb soubor s okomentovaným zdrojovým kódem.
- Výstup zašlete prosím do 14 dnů na e-mail HR specialisty, který s Vámi komunikuje.

Diskuze nad Vámi vypracovaným řešením bude součástí pohovoru. Vypracované a odevzdané řešení je podmínkou pro pokračování v přijímacím řízení.

Přejeme Vám příjemnou zábavu při řešení úloh. :)

Část 1 - aukce

Načtete soubor "bids_train.csv". Data obsahují záznamy z proběhlých aukcí. Prodejce nabízel reklamní prostor účastníkům aukce. V těchto aukcích soutěžili tři účastníci (bidders) A, B, a C, ale ne všichni se musí zúčastnit každé aukce. Hodnoty v tabulce jsou nabídky (bids) daného účastníka v dané aukci. Aukce je typu second-price; to znamená, že aukci vyhraje účastník s nejvyšší nabídkou a zaplatí druhou nejvyšší nabídku.

1. Spočítejte procenta výher pro každého účastníka. Tj. *počet aukcí, které vyhrál účastník A* děleno *celkový počet aukcí* atd.
2. Pro každou aukci určete cenu, kterou musel výherce zaplatit.

3. Jaký je celkový výnos prodejce ze všech aukcí?
4. Řekněme, že z pohledu prodejce jsou ceny náhodné veličiny. Zkuste odhadnout, jakým rozdělením pravděpodobnosti se ceny řídí.
5. Který účastník má nejvyšší ochotu platit?
6. Zobrazte vývoj ceny v čase (tj. ceny, kterou musel výherce zaplatit v každé aukci).

Část 2 - online aukce

V aukcích internetové reklamy jsou účastníci aukce inzerenti, dražený objekt je reklamní prostor a prodejcem je vlastník reklamního prostoru (například Seznam.cz). Inzerenti soutěží o zobrazení své reklamy v reklamním prostoru a platí pouze tehdy, když je jejich reklama zobrazena a návštěvník internetové stránky na ni klikne. Proto se v online aukcích výherce neurčuje pouze podle nabízené ceny, ale podle násobku nabízené ceny a odhadnuté pravděpodobnosti prokliku.

Načtete soubor “features_train.csv”, který obsahuje historická data o modelem odhadnutých pravděpodobnostech prokliku a faktorech, které tuto pravděpodobnost ovlivňují. Sloupec “click” říká, jestli daná reklama obdržela klik nebo ne. U kandidátů, kteří prohráli v aukci, je automaticky 0. Výherce aukce se potom zobrazí v reklamním prostoru, ale i tak dostane klik jen s určitou pravděpodobností.

1. Zkuste za pomoci datové analýzy odhalit a interpretovat co zachycují faktory 1 a 2.
2. Spojte tato data s daty o nabízených cenách. Pro každou aukci a každého účastníka spočítejte *online bid = nabízená cena (“bid”) * odhadnutá pravděpodobnost prokliku (“prob”)*.
3. Určete výherce každé aukce, tentokrát podle online bidu. Spočítejte, v kolika procentech případů vyhrál každý nabízející. Jak se tyto poměry změnily oproti předchozímu případu, kdy jsme uvažovali jen nabízenou cenu? Čím může být rozdíl způsoben?
4. V online aukcích se cena, kterou má zaplatit výherce aukce, počítá podle následujícího pravidla:

p.p. = odhadnutá pravděpodobnost prokliku

konečná cena za proklik = p.p. druhého v pořadí děleno p.p. prvního v pořadí krát nabízená cena druhého

Tuto cenu platí výherce pouze tehdy, pokud získal i klik. Můžete si všimnout, že finální cena nezávisí na nabízené ceně daného inzerenta.

5. Spočítejte celkový výnos z aukce.

Část 3 - model

V určitých fázích projektu může být úkolem výzkumníků vytvořit model, který bude řešit nějakou produktovou potřebu. Takový model musí být výzkumník schopen definovat, natrénovat, vyhodnotit a aplikovat.

Za použití dat ze souboru "features_train.csv":

1. Formulujte a odhadněte vlastní model, který bude predikovat pravděpodobnost, zda návštěvník internetové stránky klikne na reklamu (sloupec "click"). Reportujte údaje o modelu. Jakou lze očekávat chybu modelu?
2. Načtěte soubory "bids_test.csv" a "features_test.csv", které obsahují podobná data, nicméně tentokrát bez odhadnutých pravděpodobností prokliku.
 - 2.1. Použijte model odhadnutý v bodu 1 k predikci pravděpodobnosti prokliku.
 - 2.2. Jaký je očekávaný celkový výnos z těchto aukcí?