

Domáca úloha č. 6

2-AIN-150, Zima 2018

Termín: 8.1.2021, 23:59, boza@fmph.uniba.sk, predmet mailu: Strojove ucenie DU 6

Skôr ako sa pustíte do riešenia domácej úlohy, oboznámte sa so všeobecnými pokynmi, ktoré sú priložené na konci tohto dokumentu. Riešenia, ktoré odovzdáte, musia byť vaše vlastné. Neopisujte a nesnažte sa nájsť riešenia v literatúre alebo na internete!

Nájdenie dobrého rozdelenia

Úloha je celkom jednoduchá: Robíme regresiu. Máme jeden vstupný atribút a jeden výstupný atribút. Naprogramujte funkciu, ktorá by robila presne to, čo by sa udialo pri tréningu rozhodovacieho stromu v jednom vrchole. A zároveň sa snažte, aby táto funkcia bola, čo najefektívnejšia.

Formálne: Dané sú dvojice: $(x_1, y_1), \dots, (x_n, y_n)$. Pre dané s definujeme množiny $A_- = \{y_i | x_i < s\}$, $A_+ = \{y_i | x_i \geq s\}$ (do A_- padnú tie y_i , ktorých x_i je menšie ako s , A_+ si domyslíte). Nájdite také s , pre ktoré A_- aj A_+ sú neprázdne a navyše hodnota: $\frac{\text{Var}(A_-)|A_-| + \text{Var}(A_+)|A_+|}{n}$ je, čo najmenšia. ($\text{Var}(X) = \frac{1}{|X|} \sum_{x \in X} \left(x - \frac{1}{|X|} \sum_{y \in X} y\right)^2$, inými slovami disperzia)

V balíku k úlohe sú dva vstupy, jeden malý, druhý väčší. Boli by dobré, aby aj na tom väčšom váš program behal dostatočne rýchlo.

Vo svojich programoch môžete používať knižnice, ktoré robia základnú matiku, maticové operácie, rátajú inverzné matice, sústavy lineárnych rovníc, numericky/symbolicky derivujú. Explicitne máte zakázané použiť knižnicu scikit-learn a jej obdoby v iných jazykoch (čokoľvek, čo urobí domácu za vás na jeden riadok :).

Pokyny pre Python V balíku je súbor `template.py`, v ktorom doprogramujte funkciu `get_s(x, y)`. Program sa spúšťa príkazom `python template.py <vstupný súbor>`.

Pokyny pre iné jazyky Napíšte podobnú funkciu ako v Pythone a vhodne ju okomentujte a otestujte. Vstupný súbor obsahuje na prvom riadku číslo n . Potom nasleduje n riadkov, v každom jedno x_i y_i .

Všeobecné pokyny

Úlohy odovzdávajte mailom na mail s predmetom uvedeným v nadpise. Svoj kód vložte do prílohy mailu. **Do kódu vložte stručný komentár o vami naprogramovanej metóde.**

Ideálne odovzdávajte domáce úlohy v Pythone (doprogramujte požadované funkcionality zo zadania). Pokiaľ chcete použiť iný jazyk, môžete, ale musíte zároveň naprogramovať aj réžiu okolo (načítanie, výpis, ...). Bolo by ale vhodné, aby som váš program vedel rozbehať pod linuxom bez nutnosti inštalácie komerčných programov (t.j. overte si, či váš Matlabový kód ide spustiť v Octave, C# či funguje pod Monom, ...).