

1. Zadanie – Regresná úloha

Cieľom zadanie je navrhnúť model pre predikciu. Je možné využiť modely pre regresiu dostupné v rámci modulu scikit-learn (prípadne iných, po dohode s cvičiacim).

Maximálny počet bodov: **10** b (prípadne podľa pokynov na cvičeniach).

Podrobnejšie info na cvičeniach a na prednáške.

DÁTA:

K dispozícii sú dáta X_{public} o rozmere N vzoriek, kde každá vzorka obsahuje F príznakov, teda $N \times F$. K týmto dátam je zverejnený odpovedajúci vektor značiek y_{public} o rozmere N prvkov. Dáta je možné načítať pomocou `numpy.load()`.

Overenie vášho modelu bude realizované na dátach X_{eval} rozmeru $N_e \times F$, ktoré sú generované rovnakým kľúčom ako X_{public} .

Odobzdvávanie vektora predikovaných značiek, ktorý vygeneruje váš klasifikačný/regresný model na dátach X_{eval} je

1.12.2024 o 23:55.

HODNOTENIE:

Študent odovzdáva vektor predpovedí y_{eval} o rozmere N_e prvkov. Tento je získaný aplikovaním navrhnutého modelu na dáta X_{eval} .

Úspešnosť predikcie bude vyhodnotená metrikou **r2_score**

Na základe dosiahnutej úspešnosti pri predikcii budú študenti zoradení do tabuľky začínajúc študentom, ktorý dosiahne najvyššiu hodnotu **r2_score** na X_{eval} dátach. S klesajúcou úspešnosťou bude klesať počet bodov pridelených študentovi.

TERMÍN odovzdania:

1.12.2024 23:55

Omeškanie odovzdania zadania je penalizované odpočítaním 3b, za každý začatý deň omeškania.

ČO JE POTREBNÉ ODOVZDAŤ:

1. Vektor s predikciou o rozmere N_e prvkov, uložený pomocou príkazu `numpy.save()` v numpy formáte. Meno súboru je potrebné zadať **y_predikcia.npy** . **Nekomprimovať!!!**
2. Zdrojový kód k riešeniu (v prípade viacerých súborov zdrojového kódu, je potrebné skomprimovať do jedného balíka)
3. Krátku technickú správu **v pdf formáte** o rozsahu cca 1, maximálne 3 strany, pozostávajúcu z
 - teoretická časť - stručný teoretický opis použitého klasifikačného/regresného modelu
 - návrh riešenia - metodológia riešenia, môže obsahovať zdôvodnenie, prečo bola ktorá metóda, prípadne funkcia použitá
 - diskusia a výsledky - môže obsahovať výsledky, ktoré sa vám podarilo dosiahnuť na tréningových dátach, porovnanie rôznych prístupov ktoré ste skúšali, závery, zhodnotenie a podobne.