

# IP-feladatok futásidejének becslése

Becsó Gergely, Dankó Dorottya

2022. december 16.

## 1. A feladat

Az Adatbányászat tantárgy projektmunkájának keretében egészértékű programozási feladatok (IP = integer programming) futásidejének becslését végeztük el Python programnyelven. Ehhez regressziós modelleket és neurális hálózatokat használtunk.

### 1.1. Az adathalmaz

Az adatokat magunk generáltuk. 5000 db adott sűrűségű,  $100 \times 200$ -as ritka mátrixot generáltunk, amelyek elemei a 0 és 1 számok. Minden mátrixhoz sorsoltunk egy 200-dimenziós, 0 és 1 elemekből álló célfüggvényt is, továbbá hogy maximalizálni vagy minimalizálni szeretnénk az értéket. Az IP-feladatok tehát  $Ax \leq b$ ,  $\max / \min(cx)$  alakúak voltak, ahol  $A \in \{0, 1\}^{100 \times 200}$ ,  $b = \mathbf{1}$ ,  $c \in \{0, 1\}^{200}$ . Ezekre meghívtuk a SCIP7 IP-solvert és feljegyeztük, hogy az adott feladatokat mennyi idő alatt oldja meg és mennyi LP-feladatot old meg eközben.

## 2. A megoldás

### 2.1. Előfeldolgozás

Első lépésben a .lp formátumban lévő, IP feladatokat tartalmazó fájlokból kinyertük a számunkra fontos adatokat egy pandas dataframe-be: kétdimenziós numpy tömbök készültek a mátrixokból, egydimenziósak a célfüggvényekből. Ezeket később „kilapítottuk”: egy-egy új oszlopot hoztunk létre mind a 20000 mátrixelemnek és a 200 célfüggvényelemnek. Így létrejött az  $5000 \times 20205$  méretű dataframe, ami az adatokat tartalmazza. Ez a lépés nagy számítógépes erőforrást igényelt, a futásideje is hosszú volt, de az adathalmazt ezután le tudtuk menteni egy .csv fájlba, amit később beolvashattunk gyorsan.

### 2.2. Regressziók

Két regressziós modellt próbáltunk ki az adathalmazunkon: a lineáris regressziót és az extreme gradient boosting (XGBoost) modellt. Mivel a futásidők nagyon alacsonyak voltak (nagy részük

fél másodperc alatti) és a szórásuk is kicsi volt, elvégeztük a regressziókat a használt LP feladatok számára is, ami átlagosan 12,68 volt. Tanuló- és tesztadatokra osztottuk a dataframe-ünket. Az első 4500 oszlopon tanult, az utolsó 500 oszlopot használtuk tesztelésre.

A lineáris regresszióhoz a Scikit Learn LinearModel könyvtárát használtuk, az XGBoosthoz pedig az xgboost könyvtár XGBRegressor alkönyvtárát. Az alábbi táblázat mutatja az eredményeket. Az MSE (mean squared error) jelenti az átlagos négyzetes eltérést, az MAE (mean absolute error) az átlagos négyzetes eltérést.

	MSE (futásidő)	MAE (futásidő)	MSE (LP-k)	MAE (LP-k)
LinearRegression	0.023387	0.12330	271.9823	13.2126
XGBRegressor	0.02059	0.11538	241.16588	12.6485