

Inteligentné systémy v informatike

Technická správa

Teoretická časť

```
class sklearn.ensemble.AdaBoostRegressor(base_estimator=None, *, n_estimators=50, learning_rate=1.0, loss='linear', random_state=None)
```

AdaBoost regressor je metaodhadovač, ktorý začína nasadením regresora na pôvodnú množinu údajov a potom prispôsobí ďalšie kópie regresora na rovnakú množinu údajov, ale kde sa váhy inštancií upravujú podľa chyby aktuálnej predikcie. Následní regresisti sa preto zameriavajú skôr na ťažké prípady.

Parameters: **base_estimator***object, default=None*

Základný odhad, z ktorého je zostavený zosilnený súbor. Ak **NONE**, potom je základným odhadcom DecisionTreeRegressor (max_depth = 3)

n_estimators*int, default=50*

Maximálny počet odhadov, pri ktorých je zosilňovanie ukončené. V prípade dokonalého prispôsobenia sa proces učenia zastaví skôr.

learning_rate*float, default=1.*

Miera učenia znižuje príspevok každého regresora o learning_rate. Medzi learning_rate a n_estimators existuje kompromis.

loss*{'linear', 'square', 'exponential'}, default='linear'*

Funkcia Loss, ktorá sa má použiť pri aktualizácii váh po každom zvýšení iterácie.

random_state*int or RandomState, default=None*

Ovláda náhodné semeno dané pri každom base_estimator pri každej zvyšujúcej sa iterácii. Používa sa teda iba vtedy, keď base_estimator vystaví random_state. Okrem toho ovláda bootstrap váh použitých na tréning base_estimator pri každom zvýšení iterácie. Odošlite int pre reprodukovateľný výstup cez viaceré volania funkcií.

Návrh riešenia

Najprv podľa správnej roboty algoritmov ktore budu pracovať s datami ja perekodoval v datasete všetky data typu string v int, a potom transformoval všetky data v typ float Používaním GridSearchCV ja otestoval veľa rôznych regresívnych algoritmov z rôznymi parametrami. Pomocou porovnania výsledkov testovania ja vybral algoritmus ktorý dal najväčší score na testovacích datach. Potom tiež pomocou GridSearchCV našiel najlepšie nastavenie parametrov pre vybraný algoritmus. Podarilo sa mi dosiahnuť presnosť na testovacích datach v 40%

Diskusia a výsledky

```
AdaBoostRegressor(n_estimators=1000, random_state=8, loss='square')
```

Accuracy on training set: 0.770861

Accuracy on test set: 0.405494

--- 49.964706 seconds ---

```
GradientBoostingRegressor(loss='ls', n_estimators=1000)
```

Accuracy on training set: 0.999999

Accuracy on test set: 0.354134

--- 52.795495 seconds ---

```
RandomForestRegressor(max_depth=80, n_estimators=1000, bootstrap=True,  
random_state=0)
```

Accuracy on training set: 0.886515

Accuracy on test set: 0.195265

--- 96.402721 seconds ---