

Programowanie zaawansowane - projekt

Magdalena Czumak
Przemysław Matuszczak
Bartosz Sowiński
Norbert Sulżycki

30 czerwca 2020

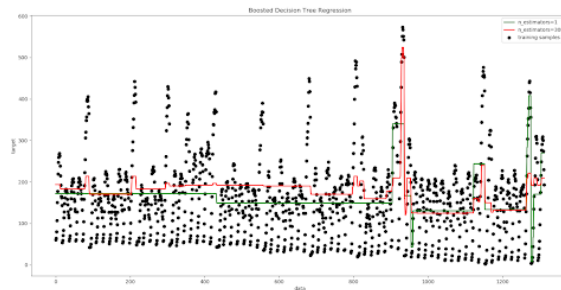
1 Problem badawczy

Problemem badawczym postawionym w niniejszej pracy jest prognoza obciążenia parkingu samochodowego umieszczonego w Birmingham, UK (regresja). Dane wykorzystane w projekcie zostały zaczerpnięte ze strony <https://archive.ics.uci.edu/ml/datasets/Parking+Birmingham> (załącznik nr 1).

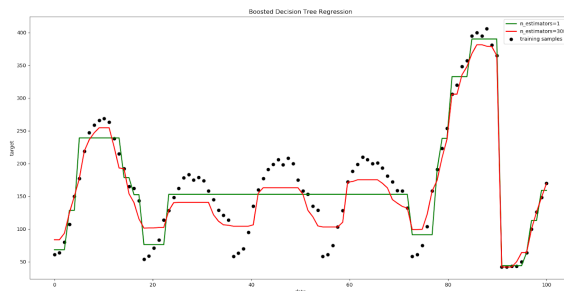
2 Model

Modele zostały zbudowane przy wykorzystaniu algorytmu Drzewa Decyzyjnego (https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html). Kod dla obu modeli jest dostępny jako załącznik nr 2.

Pierwszy wykres sporządzony został na podstawie wszystkich możliwych danych, które były dostępne. Zaobserwować można fakt, że jest on stosunkowo płaski i zdecydowanie mniej zróżnicowany niż posiadane dane (przy obu parametrach $n_{estimators}$). Parametr $n_{estimators}$ świadczy o tym, na ile części dane są dzielone, a następnie algorytm z każdej takiej pojedynczej części oblicza regresję liniową. Oznacza to, że im mniejsza wartość $n_{estimators}$, tym na mniej części dane są dzielone.



Kolejnymi zbudowanymi modelami (dla dwóch wartości $n_{estimators}$) są widoczne na poniższym wykresie linie zielona i czerwona. Zbudowane one zostały na podstawie tylko i wyłącznie dostępnych danych z sobót.



3 Analiza porównawcza

Zbudowanie powyższych modeli przy użyciu Drzewa decyzyjnego nie było trafnym pomysłem. Nie nadaje się ono do rozwiązania postawionego problemu, czyli prognozy przyszłego obciążenia parkingu. Aby możliwa była predykcja, koniecznym byłoby poznanie funkcji będącej modelem danych. W tym przypadku takich modeli jest wiele funkcji (osobno dla każdego przedziału determinowanego przez parametr $n_{estimators}$).

Jednakże warto wspomnieć inną cechę wspomnianego parametru. W przypadku mocno zróżnicowanych danych, na które wpływ miały różne szумы, zmniejszając $n_{estimators}$, zmniejszamy jednocześnie wrażliwość modelu na zakłócenia. Zjawisko to jest zdecydowanie uwidocznione w przypadku pierwszego modelu, gdzie mimo dużego zróżnicowania danych, modele są wypłaszczone.

4 Podsumowanie

Podsumowując, zastosowanie Drzewa decyzyjnego nie nadaje się do tworzenia prognozy.