

Mateusz Biegański - raport

Przetestowałem modele ElasticNet (paczka `glmnet`) oraz Random Forest (paczka `randomForest`). Jako technikę walidacji modelu użyłem sprawdzianu krzyżowego (cross-validation).

Ponieważ p jest znacznie większe od n , należało dokonać redukcji wymiaru. Jako zmienne znaczące wybrałem 500 predyktorów o największej wariancji, argumentując mój wybór tym, że zmienne o bliskiej zeru wariancji nie mogą dobrze wyjaśniać zmian zmiennej wyjaśnialnej.

Zarówno ElasticNet jak i Random Forest dały zaskakująco dobre wyniki, o błędzie MSE rzędu 5%.

Jako model ostateczny zdecydowałem się na ElasticNet z parametrem $\alpha = 0.6$, ponieważ był on obciążony najmniejszą wariancją błędów, jednocześnie dając MSE zbliżone do pozostałych.