

Porównanie błędu predykcji dla różnych metod  
estymacji współczynników w modelu liniowym  
( $p$  bliskie lub większe od  $n$ ) - studium symulacyjne.

**Przemysław Biecek**

Zakład Statystyki  
Instytut Matematyki Stosowanej  
Uniwersytet Warszawski  
Banacha 2, 02-097 Warszawa  
e-mail: przemyslaw.biecek@gmail.com

**Agnieszka Prochenka**

Uniwersytet Warszawski  
Banacha 2, 02-097 Warszawa

W wielu praktycznych zastosowaniach model liniowy wykorzystywany jest do predykcji zmiennej objaśnianej. Taką predykcję można wykonać na wiele rozmaitych sposobów: estymując wszystkie współczynniki w pełnym modelu liniowym, estymując zbiór istotnych współczynników (istotnych w sensie pewnego kryterium, np. BIC, AIC, FDR lub innego), korzystając z regresji grzbietowej, regresji lasso lub innych metod predykcji.

Naturalnym pytaniem jest jak zachowuje się błąd predykcji dla różnych metod. Ponieważ dla wielu metod trudno jest badać analitycznie błąd predykcji przeprowadziliśmy studium symulacyjne.

Podczas referatu przedstawimy pokrótce wybrane metody, które można wykorzystać do predykcji modelu, porównamy zdolności predykcyjne modeli zbudowanych w oparciu o poszczególne metody oraz przedyskutujemy otrzymane wyniki.

Sz szczególnie interesować będzie nas przypadek w którym kąt pomiędzy kolumnami macierzy eksperymentu jest (bardzo) ostry oraz przypadek gdy liczba zmiennych  $p$  jest duża, tj. porównywalna lub większa od  $n$ . Przedstawimy też przykładowe problemy dotyczących rzeczywistych analiz w których mamy do czynienia z takimi scenariuszami.