

# Modelowanie Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

## Laboratorium 9.

Na podstawie zbioru danych aplikacyjnych oraz behawioralnych dostępnych w pliku `KrukUWr2018.RData` wykonaj poniższe polecenia.

1. Przygotuj dane aplikacyjne do analizy.
2. Zadaniem jest wybór optymalnego zbioru cech objaśniających w liniowym modelu regresji modelującym SR12M. W tym celu stwórz wykres obrazujący wszystkie możliwe kombinacje cech objaśniających oraz RSS (residual sum of squares) modelu obliczony na podstawie próby testowej (oś odciętych to liczba cech objaśniających, zaś oś rzędnych to RSS).
  - Jaka jest liczba wszystkich możliwych kombinacji cech objaśniających, jeżeli do dyspozycji jest  $p$  cech?
  - Jaka jest optymalna kombinacja cech?
  - Dokonaj interpretacji oszacowanych parametrów wybranej kombinacji.
3. Jaka jest kombinacja cech objaśniających wskazywana przez *forward* oraz *backward stepwise selection*? Jak wyglądają ścieżki tych dwóch podejść na wykresie z zadania 2?
4. Oszacuj liniowy model regresji grzbietowej na wyjściowym zbiorze cech objaśniających dla różnych poziomów parametru regularizacyjnego  $\lambda$ :

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}, \text{ gdzie } \lambda \geq 0.$$

- Narysuj wykres oszacowania błędu testowego w zależności od parametru  $\lambda$ .
  - Jaka jest optymalna wartość parametru  $\lambda$ ?
5. Oszacuj liniowy model regresji z regularyzacją *lasso* na wyjściowym zbiorze cech objaśniających dla różnych poziomów parametru regularizacyjnego  $\lambda$ :

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \text{ gdzie } \lambda \geq 0.$$

- Narysuj wykres oszacowania błędu testowego w zależności od parametru  $\lambda$ .
- Jaka jest optymalna wartość parametru  $\lambda$ ?

Listę zadań podsumuj wypełniając następującą tabelę:

Cecha	Wszystkie cechy	Najlepszy podzbiór	Ridge	Lasso	Forward stepwise selection	Backward stepwise selection
Cecha 1						
⋮				(oszacowanie parametru)		
Cecha $p$						
Oszacowanie błędu testowego						