

# Laboratoria 2

Agata Czajkowska, Artur Minorczyk

Październik 16, 2017

## 1 . Otrzymane wyniki

W projekcie należało dokonać zadania predykcji na sztucznie wygenerowanych danych. Dane składały się z 50 cech i jednej zmiennej zależnej, która mogła przyjąć 2 możliwe wartości “klasa +” lub “klasa -”. By w efekcie możliwe było sprawdzenie jakości predykcji dostarczony zbiór treningowy podzielono w sposób losowy na zbiór treningowy i testowy z zachowaniem proporcji 4:1. Wszystkie testy, decyzje o istotności zmiennych wykonano na poziomie istotności  $\alpha = 0.05$ . Otrzymano następujące wyniki.

Metoda	Model	Wynik
Lasy losowe	Krokowy	0.8260
Lasy losowe	Regresja logistyczna	0.8055
Regresja logistyczna	Krokowy	0.7195
Regresja logistyczna	Regresja logistyczna	0.7160

## 2. Użyte metody

W projekcie użyto 2 rodzajów klasyfikacji : **regresji logistycznej** i **algorytmu lasów losowych** . Wcześniej dokonano selekcji zmiennych na każdy z dwóch sposobów : **procedury krokowej** ( z kryterium AIC i kierunkiem wstecz), oraz **regresji logistycznej**. W efekcie przetestowano 4 przypadki , na obu modelach powstałych po selekcji zmiennych sprawdzono każdy z algorytmów klasyfikacji.

## 3. Wykres

## 4. Poprawność otrzymanego wyniku

Jakość klasyfikacji obliczono za pomocą wzoru  $wynik = \frac{\sum_{i=1}^{0.2*|X_{test}|} I(x_i)}{0.2*|X_{test}|}$  , gdzie  $I(x) = \begin{cases} 1, & \text{if } x \in \text{klasa+} \\ 0, & \text{if } x \notin \text{klasa+} \end{cases}$  ,  $X_{test}$  to zbiór obserwacji testowych uszeregowany nierosnąco względem predykcji przynależności do klasy + .