

# Klasyfikacja i regresja: Implementacja naiwnego klasyfikatora bayesowskiego

7 listopada 2018

## Opis pliku z zadaniami

Wszystkie zadania na zajęciach będą przekazywane w postaci plików `.pdf`, sformatowanych podobnie do tego dokumentu. Zadania będą różnego rodzaju. Za każdym razem będą one odpowiednio oznaczone:

- Zadania do wykonania na zajęciach oznaczone są symbolem  $\triangle$  – nie są one punktowane, ale należy je wykonać w czasie zajęć.
- Punktowane zadania do wykonania na zajęciach oznaczone są symbolem  $\diamond$  – należy je wykonać na zajęciach i zaprezentować prowadzącemu.
- Zadania do wykonania w domu oznaczone są symbolem  $\star$  – są one punktowane, należy je dostarczyć w sposób podany przez prowadzącego i w wyznaczonym terminie (zwykle przed kolejnymi zajęciami).

# 1 Implementacja naiwnego klasyfikatora Bayesa dla danych numerycznych oraz nominalnych 10p.◇

## Treść

Zmodyfikuj poprzednią implementację algorytmu naiwnego klasyfikatora Bayesa z poprzednich zajęć, tak aby obsługiwać także dane numeryczne. Można wykorzystać ten sam dostarczony szkielet oprogramowania.

Szczegóły zadania:

- Obsłuż cechy numeryczne.
- Przetestuj swoją implementację przy użyciu zbioru danych Iris dostarczonego w pakiecie sklearn.

Kilka podpowiedzi:

- Załóż, że dane numeryczne pochodzą z rozkładu normalnego dla każdej klasy. Funkcja gęstości prawdopodobieństwa rozkładu normalnego wyrażona jest wzorem:

$$g_y(x) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_y)^2}{2\sigma_y^2}\right),$$

gdzie  $\mu_y$  i  $\sigma_y$  to odpowiednio średnia i odchylenie standardowe obserwacji na danej cesze dla klasy  $y$ .

- Przechowuj w odpowiednich strukturach danych średnie i odchylenia standardowe dla atrybutów numerycznych (dla każdej klasy osobno) oraz prawdopodobieństwa a priori dla poszczególnych klas.