

STATYSTYCZNE METODY REGRESJI PORZĄDKOWEJ

Marta Sommer

MiNI, Politechnika Warszawska

25 stycznia 2016 r.



MOTYWACJA

SZUKANE

ZWYKŁA KLASYFIKACJA

Czy klientowi
spodoba się dana
książka?

- 1) nie
- 2) tak

REGRESJA PORZĄDKOWA

W jakim stopniu klientowi
spodoba się dana książka?

- 1) bardzo się nie spodoba
- 2) raczej się nie spodoba
- 3) ani się spodoba ani się nie spodoba
- 4) raczej się spodoba
- 5) bardzo się spodoba

DANE

Co wiemy o
klientcie?

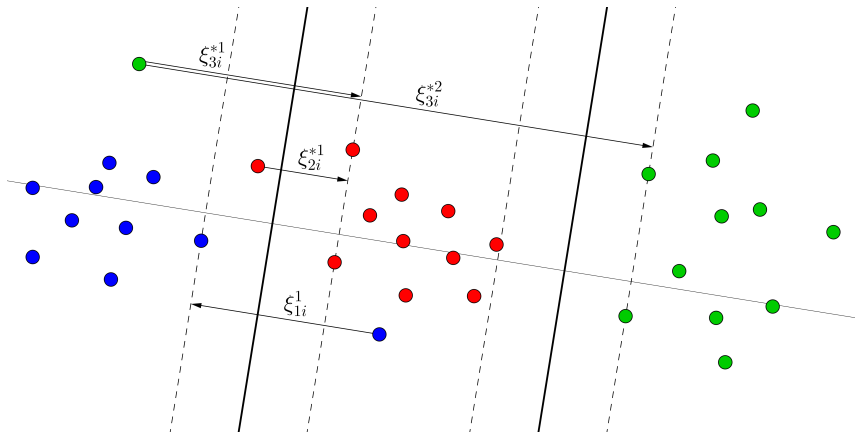
- wiek
- płeć
- wykształcenie
- historia zakupów
- ...

DOSTĘPNE METODY

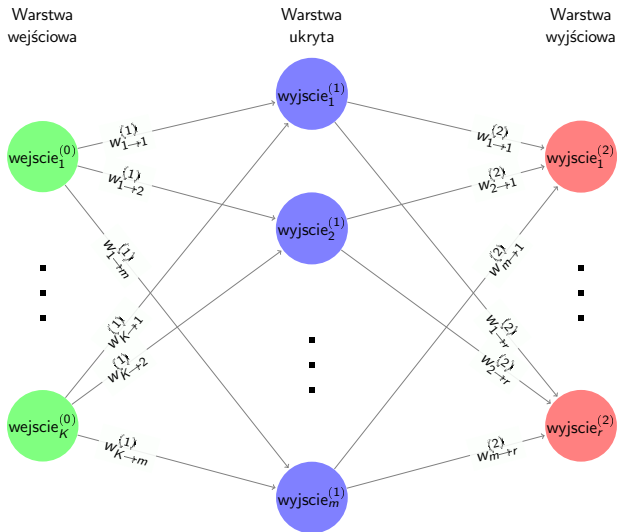
- model proporcjonalnych szans

$$\log \frac{\mathbb{P}(Y \leq j \mid \mathbf{x})}{1 - \mathbb{P}(Y \leq j \mid \mathbf{x})} = \alpha_j + \beta^T \mathbf{x}$$

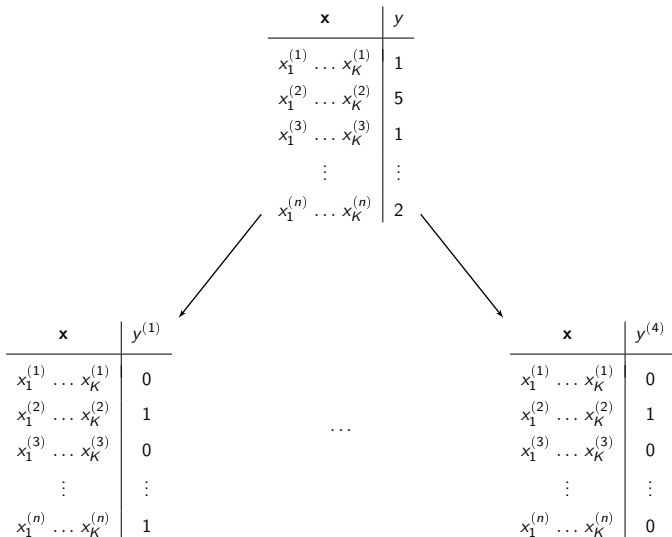
- maszyna wektorów podpierających (SVM)



- sieci neuronowe



- metoda Franka i Halla



DIAGNOSTYKA MODELU

- procent poprawnej klasyfikacji

$$ACC = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{y_i = \hat{y}_i\}$$

- średni błąd bezwzględny

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- współczynnik VUS

$$VUS = \frac{1}{n_1 n_2 \cdot \dots \cdot n_r} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_r=1}^{n_r} \mathbb{I}\{f(\mathbf{x}_{i_1}^1) < \dots < f(\mathbf{x}_{i_r}^r)\}$$

- współczynnik BSC (oparty na sortowaniu bąbelkowym)