

STATYSTYCZNE METODY REGRESJI PORZĄDKOWEJ

Marta Sommer

MiNI, Politechnika Warszawska

10 marca 2015



PRZYKŁAD

Założmy, że chcemy modelować zmienną odpowiedzi przyjmującą wartości ze zbioru:

$$\mathcal{Y} = \{ \textit{bardzo_sie_nie_zgadzam}, \textit{nie_zgadzam_sie}, \\ \textit{zgadzam_sie}, \textit{bardzo_sie_zgadzam} \}.$$

W naturalny sposób są to dane uporządkowane:

$$\textit{bardzo_sie_nie_zgadzam} < \textit{nie_zgadzam_sie} < \\ < \textit{zgadzam_sie} < \textit{bardzo_sie_zgadzam}.$$

Mamy też dane zmienne objaśniające - oznaczmy je standardowo przez \mathcal{X} .

MODEL PROPORCJONALNYCH SZANS

Interesują nas:

$$\Pi_j(\underline{x}) = \mathbb{P}(Y = j \mid \underline{x}), \quad \text{dla } j = 1, \dots, J.$$

Rozważamy skumulowane prawdopodobieństwa:

$$\mathbb{P}(Y \leq j \mid \underline{x}) = \Pi_1(\underline{x}) + \dots + \Pi_j(\underline{x}).$$

Oraz model:

$$\log \frac{\mathbb{P}(Y \leq j \mid \underline{x})}{\mathbb{P}(Y > j \mid \underline{x})} = \alpha_j + \underline{\beta}' \underline{x}.$$

Współczynniki wyliczamy metodą Raphsona-Newtona, a szukane prawdopodobieństwa - po prostym przeliczeniu - dostaniemy ze wzoru:

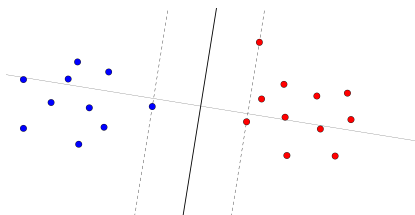
$$\mathbb{P}(Y \leq j \mid \underline{x}) = \frac{e^{\alpha_j + \underline{\beta}' \underline{x}}}{1 + e^{\alpha_j + \underline{\beta}' \underline{x}}}.$$

WEKTORY MASZYN PODPIERAJĄCYCH (SVM)

PRZYPADEK DWUKLASOWY, LINIOWO SEPAROWALNY

Konstruujemy dwie równoległe i maksymalnie oddalone od siebie hiperpłaszczyzny, we wnętrzu których nie leży ani jeden element próby uczącej. Ich równania to:

$$\begin{cases} x_i' \mathbf{w} + b = 1 \\ x_i' \mathbf{w} + b = -1 \end{cases}$$



Odległość między hiperpłaszczyznami wynosi: $\frac{2}{\|\mathbf{w}\|}$, więc celem jest minimalizacja $\frac{\|\mathbf{w}\|^2}{2}$ przy ograniczeniach:

$$\begin{cases} x_i' \mathbf{w} + b \geq 1 \\ x_i' \mathbf{w} + b \leq -1 \end{cases}$$

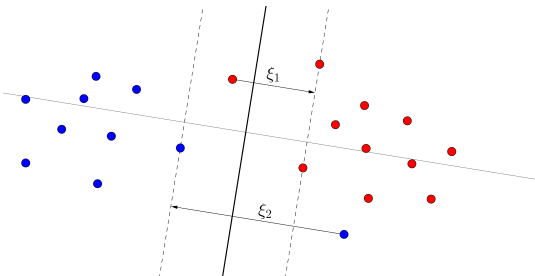
PRZYPADEK DWUKLASOWY, LINIOWO NIESEPAROWALNY

Dokładamy stałe $\xi_i \geq 0$ osłabiające warunek liniowej separowalności (kary za nieidealne rozdzielenie). W takiej sytuacji rozwiązujemy problem optymalizacyjny:

$$\min \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right\}$$

przy ograniczeniach:

$$\begin{cases} x_i' \mathbf{w} + b \geq 1 - \xi_i \\ x_i' \mathbf{w} + b \leq -1 + \xi_i \end{cases}$$



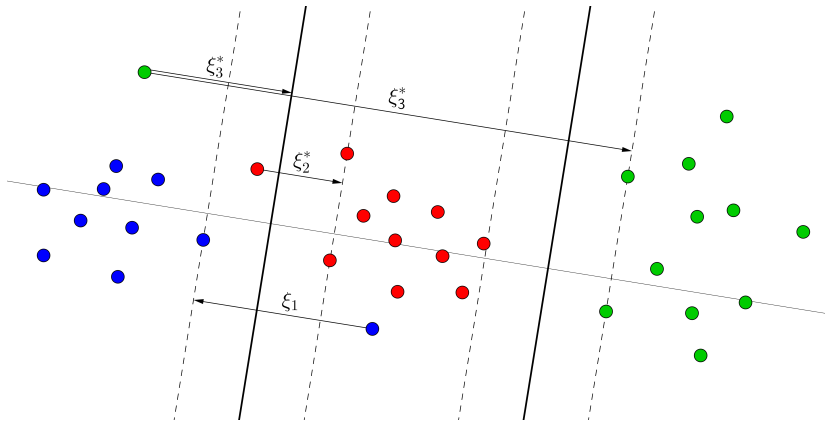
PRZYPADEK J UPORZĄDKOWANYCH KLAS

Optymalizujemy:

$$\min_{\mathbf{w}, b_k} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^{J-1} \left(\sum_{k=1}^j \sum_{i=1}^{n_k} \xi_{ki}^j + \sum_{k=1}^j \sum_{i=1}^{n_k} \xi_{ki}^{*j} \right) \right\}$$

przy ograniczeniach:

$$\begin{cases} \mathbf{w}x_i^k - b_j \leq -1 + \xi_{ki}^j, & \text{dla } k = 1, \dots, j \text{ oraz } i = 1, \dots, n_k \\ \mathbf{w}x_i^k - b_j \geq +1 - \xi_{ki}^{*j}, & \text{dla } k = j+1, \dots, J \text{ oraz } i = 1, \dots, n_k \end{cases}$$



**METODA
ZAPROPONOWANA
PRZEZ E. FRANKA
I M. HALLA**

Podejście Franka i Halla do zagadnienia regresji porządkowej opiera się nie na stworzeniu nowego modelu, ale na odpowiednim przedefiniowaniu zbioru danych, a następnie na sprowadzeniu zadania do problemu zwykłej klasyfikacji z dwoma klasami.

J –klasowy problem
regresji porządkowej



$J - 1$ dwuklasowych
problemów klasyfikacji

Na czym więc polega transformacja danych? Przeanalizujmy to na przykładzie:

$$\mathcal{Y} = \{ \textit{bardzo_sie_nie_zgadzam}, \quad \textit{nie_zgadzam_sie}, \\ \textit{zgadzam_sie}, \quad \textit{bardzo_sie_zgadzam} \}.$$

Dzielimy 4–klasowy problem regresji porządkowej na 3 dwuklasowe problemy klasyfikacji w następujący sposób:

Zbiór nr 1: $Cel > \textit{bardzo_sie_nie_zgadzam}$

Zbiór nr 2: $Cel > \textit{nie_zgadzam_sie}$

Zbiór nr 3: $Cel > \textit{zgadzam_sie}$

Przy czym modyfikacji ulega jedynie wektor \mathcal{Y} , zbiór \mathcal{X} pozostaje bez zmian.

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>

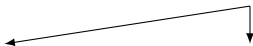
\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	\mathcal{Y}
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 2:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 2:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 3:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	0
1 6 5 B	0

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 2:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 3:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	0
1 6 5 B	0



$$\mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} | \underline{x})$$

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 2:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 3:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	0
1 6 5 B	0



$$\mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} | \underline{x})$$



$$\mathbb{P}(Cel > \textit{nie_zgadzam_sie} | \underline{x})$$

\mathcal{X}	\mathcal{Y}
1 3 5 A	<i>bardzo_sie_nie_zgadzam</i>
3 3 3 A	<i>bardzo_sie_nie_zgadzam</i>
1 4 4 F	<i>nie_zgadzam_sie</i>
2 3 3 A	<i>bardzo_sie_zgadzam</i>
5 9 1 A	<i>zgadzam_sie</i>
1 6 5 B	<i>bardzo_sie_nie_zgadzam</i>



Zbiór nr 1:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	1
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 2:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	1
1 6 5 B	0

Zbiór nr 3:

\mathcal{X}	Cel
1 3 5 A	0
3 3 3 A	0
1 4 4 F	0
2 3 3 A	1
5 9 1 A	0
1 6 5 B	0



$$\mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} | \underline{x})$$



$$\mathbb{P}(Cel > \textit{nie_zgadzam_sie} | \underline{x})$$



$$\mathbb{P}(Cel > \textit{zgadzam_sie} | \underline{x})$$

Następnie dostajemy nową obserwację o zmiennych objaśnianych \underline{x} i chcemy dowiedzieć się, do jakiej klasy z \mathcal{Y} będzie ona należała. Co robimy? Korzystając z poprzednich modeli, liczymy po kolei:

$$\mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x})$$

$$\mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x})$$

$$\mathbb{P}(Cel > \textit{zgadzam_sie} \mid \underline{x})$$

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

$$\mathbb{P}(Cel = \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) = 1 - \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x})$$

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

$$\begin{aligned}\mathbb{P}(Cel = \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) &= 1 - \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) \\ \mathbb{P}(Cel = \textit{nie_zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x})\end{aligned}$$

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

$$\mathbb{P}(Cel = \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) = 1 - \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x})$$

$$\begin{aligned} \mathbb{P}(Cel = \textit{nie_zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x}) \end{aligned}$$

$$\begin{aligned} \mathbb{P}(Cel = \textit{zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{zgadzam_sie} \mid \underline{x}) \end{aligned}$$

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

$$\mathbb{P}(Cel = \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) = 1 - \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x})$$

$$\begin{aligned} \mathbb{P}(Cel = \textit{nie_zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x}) \end{aligned}$$

$$\begin{aligned} \mathbb{P}(Cel = \textit{zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{zgadzam_sie} \mid \underline{x}) \end{aligned}$$

$$\mathbb{P}(Cel = \textit{bardzo_sie_zgadzam} \mid \underline{x}) = \mathbb{P}(Cel > \textit{bardzo_sie_zgadzam} \mid \underline{x})$$

Kolejne prawdopodobieństwa wylicza się łańcuchowo. Mianowicie:

$$\mathbb{P}(Cel = \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) = 1 - \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x})$$

$$\begin{aligned}\mathbb{P}(Cel = \textit{nie_zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{bardzo_sie_nie_zgadzam} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x})\end{aligned}$$

$$\begin{aligned}\mathbb{P}(Cel = \textit{zgadzam_sie} \mid \underline{x}) &= \mathbb{P}(Cel > \textit{nie_zgadzam_sie} \mid \underline{x}) - \\ &\quad \mathbb{P}(Cel > \textit{zgadzam_sie} \mid \underline{x})\end{aligned}$$

$$\mathbb{P}(Cel = \textit{bardzo_sie_zgadzam} \mid \underline{x}) = \mathbb{P}(Cel > \textit{bardzo_sie_zgadzam} \mid \underline{x})$$

Ostatecznie, naszej nowej obserwacji przypisujemy klasę z maksymalnym prawdopodobieństwem.

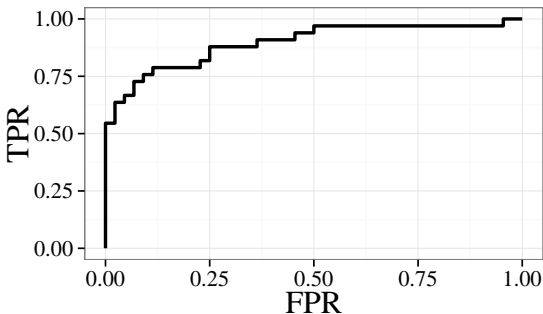
KRZYWA ROC I WSPÓŁCZYNNIK AUC

PRZYPADEK DWUKLASOWY

		Klasa	
		1	0
\widehat{Klasa}	1	TP	FP
	0	FN	TN

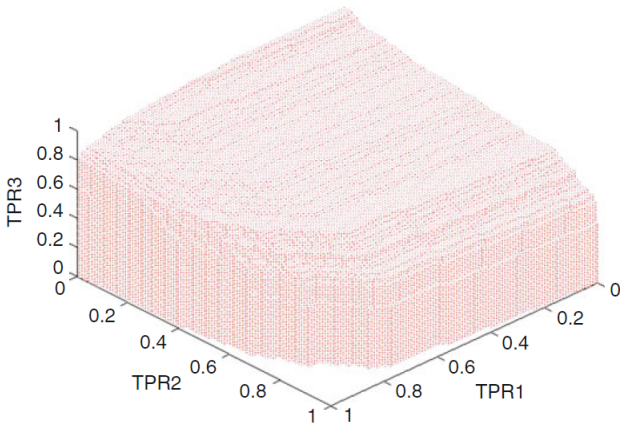
$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$



Wykres 1: Przykładowa krzywa ROC

PRZYPADEK KLASYFIKACJI PORZĄDKOWEJ



$$VUS = \frac{1}{\prod_{k=1}^J n_k} \sum_{y_{j_1} < \dots < y_{j_J}} \mathbb{I}_{f(x_{j_1}) < \dots < f(x_{j_J})}$$

Bibliografia



Frank E., Hall M., A simple approach to ordinal classification, *Proceedings of the European Conference on Machine Learning*, Freiburg, Niemcy, 2001, str. 146–156.



Waegman W., De Baets B., A survey on ROC-based ordinal regression, w: Fürnkranz J., Hüllermeier E. (Eds.), *Preference Learning*, Springer, 2010, str. 127-154.