UNIWERSYTET
MIKOŁAJA KOPERNIKA
W TORUNIU

Możliwości predykcyjne krzywych ROC

Ośrodek Analiz Statystycznych UMK

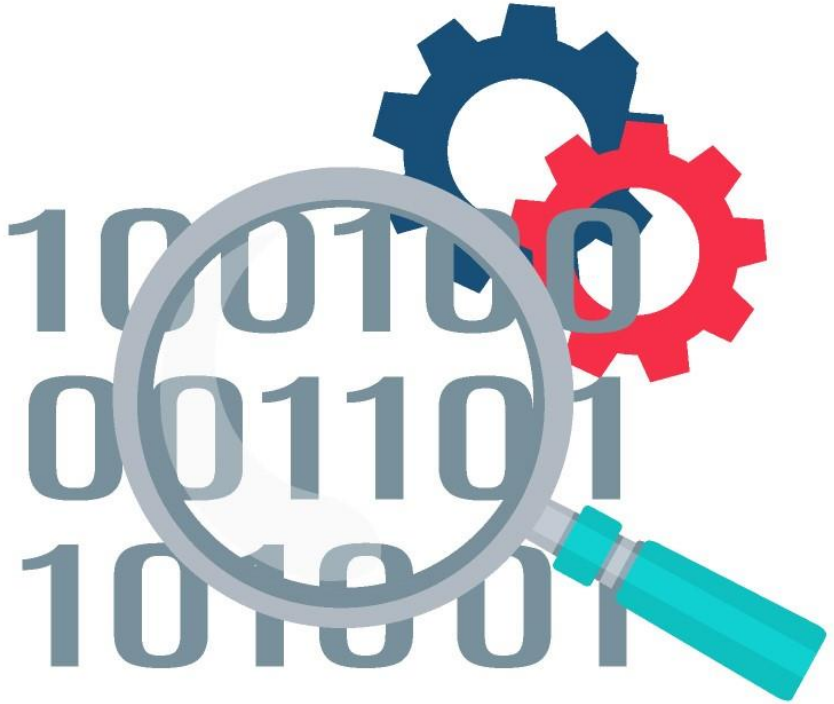
dr Joanna Karłowska-Pik, inż. Krzysztof Leki

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Spis treści

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

1. JAKOŚĆ KLASYFIKACJI BINARNEJ



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Klasyfikacja binarna

X_1	X_2	...	X_p	faktyczne Y
x_{11}	x_{12}	...	x_{1p}	0
x_{21}	x_{22}	...	x_{2p}	1
x_{31}	x_{32}	...	x_{3p}	1
x_{41}	x_{42}	...	x_{4p}	0
⋮	⋮	⋮	⋮	⋮

W przypadku klasyfikacji binarnej zmienna celu Y ma dwie wartości, najczęściej oznaczane jako 0 (negatywne) i 1 (pozytywne), przy czym 1 nie musi być pozytywna w wydźwięku – z reguły oznacza zdarzenie, które nas bardziej interesuje (np. chorobę).

Modele predykcyjne

X_1	X_2	...	X_p	faktyczne Y	$P(Y = 1 \mid X_1, \dots, X_p)$
x_{11}	x_{12}	...	x_{1p}	0	p_1
x_{21}	x_{22}	...	x_{2p}	1	p_2
x_{31}	x_{32}	...	x_{3p}	1	p_3
x_{41}	x_{42}	...	x_{4p}	0	p_4
⋮	⋮	⋮	⋮	⋮	⋮

Gdy budujemy modele predykcyjne, na podstawie wartości dostępnych zmiennych (predyktorów) $X_1, X_2, \dots, X_p$ określane jest prawdopodobieństwo, że wartością zmiennej Y jest 1.

Predykcja w oparciu o prawdopodobieństwo

X_1	X_2	...	X_p	faktyczne Y	$P(Y = 1 \mid X_1, \ldots, X_p)$	przewidywane $\widehat{Y}$
x_{11}	x_{12}	...	x_{1p}	0	p_1	0
x_{21}	x_{22}	...	x_{2p}	1	p_2	1
x_{31}	x_{32}	...	x_{3p}	1	p_3	0
x_{41}	x_{42}	...	x_{4p}	0	p_4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Dla określonego progu prawdopodobieństwa możemy dokonać predykcji wartości zmiennej Y, przyjmując 1, gdy wartość prawdopodobieństwa przekracza wartość progową, a 0 w przeciwnym wypadku.

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Predykcja w oparciu o wybrany predyktor

X_1	X_2	...	X_p	faktyczne Y	$P(Y=1\|X_1,\dots,X_p)$	przewidywane $\widehat{Y}$
x_{11}	x_{12}	...	x_{1p}	0	p_1	0
x_{21}	x_{22}	...	x_{2p}	1	p_2	1
x_{31}	x_{32}	...	x_{3p}	1	p_3	0
x_{41}	x_{42}	...	x_{4p}	0	p_4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Podobnie możemy postąpić, posługując się zamiast prawdopodobieństwami wartościami dowolnego predyktora (najczęściej biomarkera, zmiennej o właściwościach diagnostycznych). Należy jednak zwrócić uwagę, czy na korzyść faktu, że $Y = 1$, świadczą coraz większe, czy coraz mniejsze wartości rozważanego predyktora.

W dalszych rozważaniach nie będziemy odróżniać, czy predykcji dokonujemy w oparciu o model oparty o wiele predyktorów i uzyskane z niego prawdopodobieństwa, czy o pojedynczy predyktor. I prawdopodobieństwa, i pojedynczy predyktor będziemy oznaczać przez X i będziemy przyjmować, że duże wartości X świadczą na korzyść zdarzenia $Y = 1$, a małe na korzyść $Y = 0$.

Zatem

$$\hat{Y} = 1 \Leftrightarrow X \geq x$$

Rodzaje klasyfikacji

X_1	X_2	...	X_p	faktyczne Y	$P(Y = 1 \mid X_1, \ldots, X_p)$	przewidywane $\widehat{Y}$	Klasyfikacja
x_{11}	x_{12}	...	x_{1p}	0	p_1	0	TN
x_{21}	x_{22}	...	x_{2p}	1	p_2	1	TP
x_{31}	x_{32}	...	x_{3p}	1	p_3	0	FN
x_{41}	x_{42}	...	x_{4p}	0	p_4	1	FP
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Możemy porównać faktyczne i przewidywane wartości zmiennej celu, co pozwala na określenie rodzaju klasyfikacji.

- **Klasyfikacje prawdziwie negatywne** (ang. *true negatives*, TN) to rekordy, które zostały sklasyfikowane jako negatywne i w rzeczywistości są negatywne (np. zdrowy jako zdrowy).
- **Klasyfikacje prawdziwie pozytywne** (ang. *true positives*, TP) to rekordy, które zostały sklasyfikowane jako pozytywne i w rzeczywistości są pozytywne (np. chory jako chory).
- **Klasyfikacje fałszywie negatywne** (ang. *false negatives*, FN) to rekordy, które zostały sklasyfikowane jako negatywne, gdy w rzeczywistości są pozytywne (chory jako zdrowy).
- **Klasyfikacje fałszywie pozytywne** (ang. *false positives*, FP) to rekordy, które zostały sklasyfikowane jako pozytywne, gdy w rzeczywistości są negatywne (zdrowy jako chory).

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Macierz pomyłek

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Przykład:

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

Zliczamy ile jest każdego rodzaju klasyfikacji i liczebności te zamieszczamy w tabeli nazywanej **macierzą pomyłek**.

W 40 przypadkach zmienna Y miała wartość 0 i model poprawnie przewidział 0, a w 10 przypadkach pomylił się i przewidział 1. W 20 przypadkach zmienna Y miała wartość 1, ale model pomylił się i przewidział 0, zaś w 30 przypadkach poprawnie przewidział 1.

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Trafność

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Trafność (ang. *accuracy*) określa jaki odsetek rekordów ma poprawnie przewidzianą wartość zmiennej celu.

$$\frac{TN + TP}{TN + FP + FN + TP}$$

Przykład:

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

$$\frac{40 + 30}{40 + 10 + 20 + 30} = 70\%$$

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Całkowity współczynnik błędu

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Całkowity współczynnik błędu (ang. *overall error rate*) – odsetek błędnych klasyfikacji, czyli 1 – trafność.

$$\frac{FN + FP}{TN + FP + FN + TP}$$

Przykład:

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

$$\frac{20 + 10}{40 + 10 + 20 + 30} = 30\%$$

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Jaka musi być trafność, żeby model był dobry?

Przypuśćmy, że mamy w zbiorze 90 rekordów negatywnych i 10 pozytywnych. Możemy zbudować klasyfikator w oparciu o rzut monetą – wtedy będzie on miał trafność 50%. Możemy też po prostu wszystko klasyfikować do klasy negatywnej – częstszej.

		wynik	
		nie	tak
choroba	nie	90	0
	tak	10	0

Trafność takiego klasyfikatora będzie wynosić aż 90%, a on wcale nie jest dobry (nie wykrywa żadnych obserwacji pozytywnych).

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

		wynik	
		nie	tak
choroba	nie	90	0
	tak	10	0

		wynik	
		nie	tak
choroba	nie	75	15
	tak	0	10

Dla macierzy pomyłek znajdującej się po prawej stronie trafność jest niższa (85%), ale otrzymane wyniki świadczą o lepszej jakości modelu.

Wniosek: Trafność nie jest wystarczającą miarą do oceny jakości klasyfikacji.

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Czułość

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Czułość, skuteczność (ang. *sensitivity, recall, true positive rate TPR*) – jaka jest skuteczność modelu przy przewidywaniu klasy 1.

$$\frac{TP}{FN + TP} \approx P(X \geq x | Y = 1)$$

Przykład:

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

$$\frac{30}{20 + 30} = 60\%$$

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Specyficzność

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Specyficzność, swoistość (ang. *specifity, true negative rate TNR*) – jaka jest skuteczność modelu przy przewidywaniu klasy 0.

$$\frac{TN}{TN + FP} \approx P(X < x | Y = 0)$$

Przykład:

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

$$\frac{40}{40 + 10} = 80\%$$

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Wskaźnik fałszywie pozytywnych

		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	TN	FP
	1	FN	TP

Wskaźnik fałszywie pozytywnych (ang. *false positive rate, FPR*) – odsetek błędnie sklasyfikowanych przypadków negatywnych, czyli 1 – specyficzność.

Przykład:

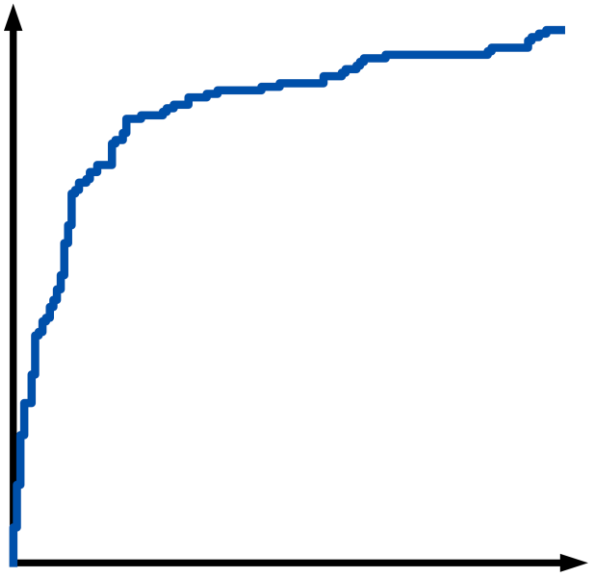
		przewidywane $\widehat{Y}$	
		0	**1**
faktyczne Y	**0**	40	10
	1	20	30

$$\frac{FP}{TN + FP} \approx P(X \geq x | Y = 0)$$

$$\frac{10}{40 + 10} = 20\%$$

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

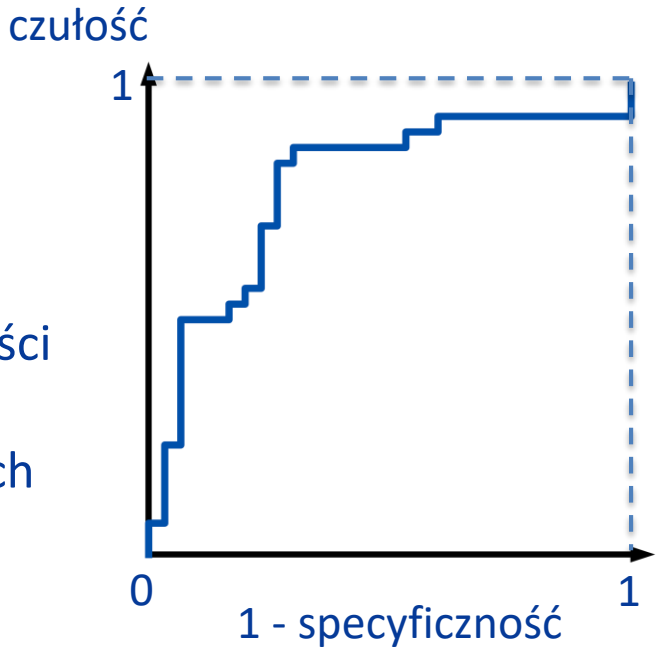
2.
KRZYWA ROC



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Empiryczna krzywa ROC

- **Krzywa ROC** – krzywa operacyjno-charakterystyczna odbiornika (ang. *Receiver Operating Characteristic curve*) – krzywa parametryzowana zmienną ciągłą (prawdopodobieństwem lub wybranym predyktorem X), wykreślająca wskaźnik czułości w stosunku do wskaźnika fałszywie pozytywnych (1 – specyficzność) dla wszystkich możliwych punktów odcięcia.
- Krzywą ROC można wyznaczyć tylko, gdy w próbce każda obserwacja ma znane wartości zmiennej celu oraz predyktora.

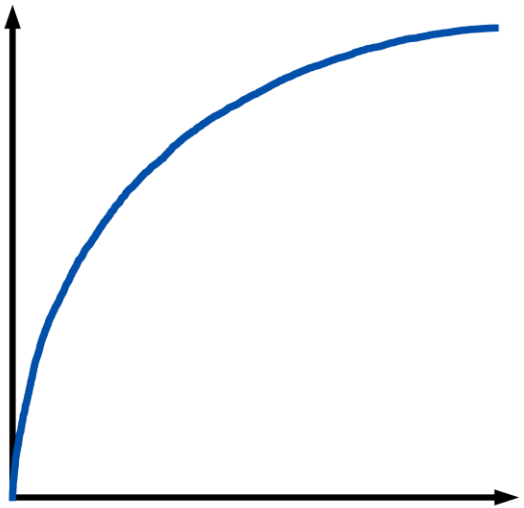


Kroki, które należy wykonać, żeby wykreślić krzywą:

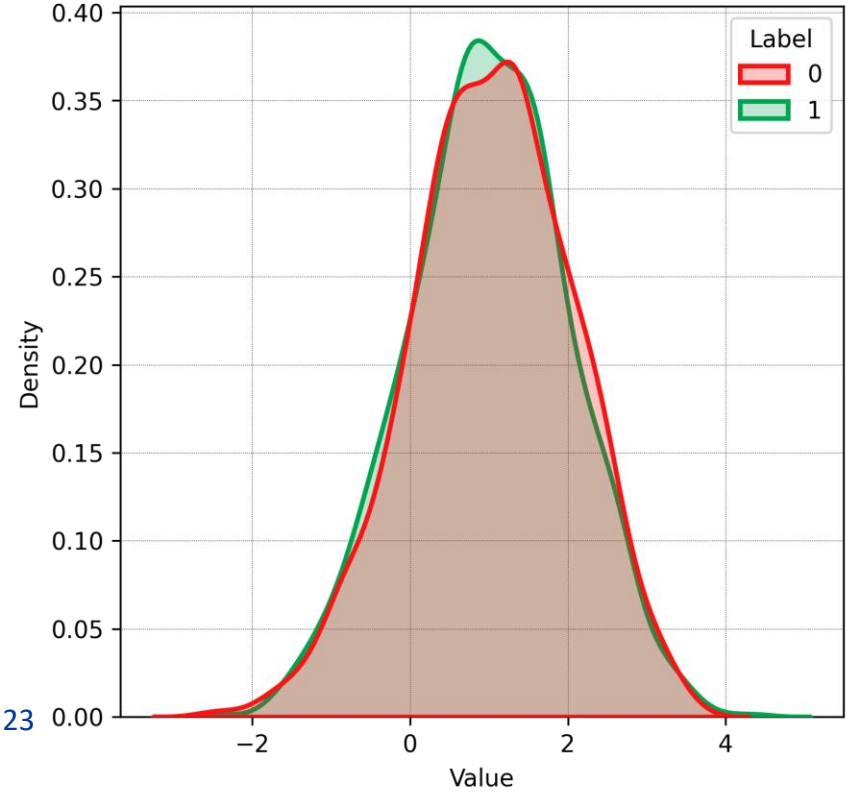
1. Sortujemy rekordy malejąco ze względu na wartości zmiennej ilościowej X.
2. Każda z tych wartości staje się kolejno progiem powyżej którego (lub poniżej którego, gdy o pozytywnej wartości zmiennej celu świadczą małe a nie duże wartości predyktora) klasyfikujemy zmienną jako 1,
   a w przeciwnym wypadku jako 0.
3. Dla każdego progu wyznaczamy czułość oraz specyficzność.
4. Zaznaczamy na wykresie punkt o współrzędnych
   (1−specyficzność, czułość).

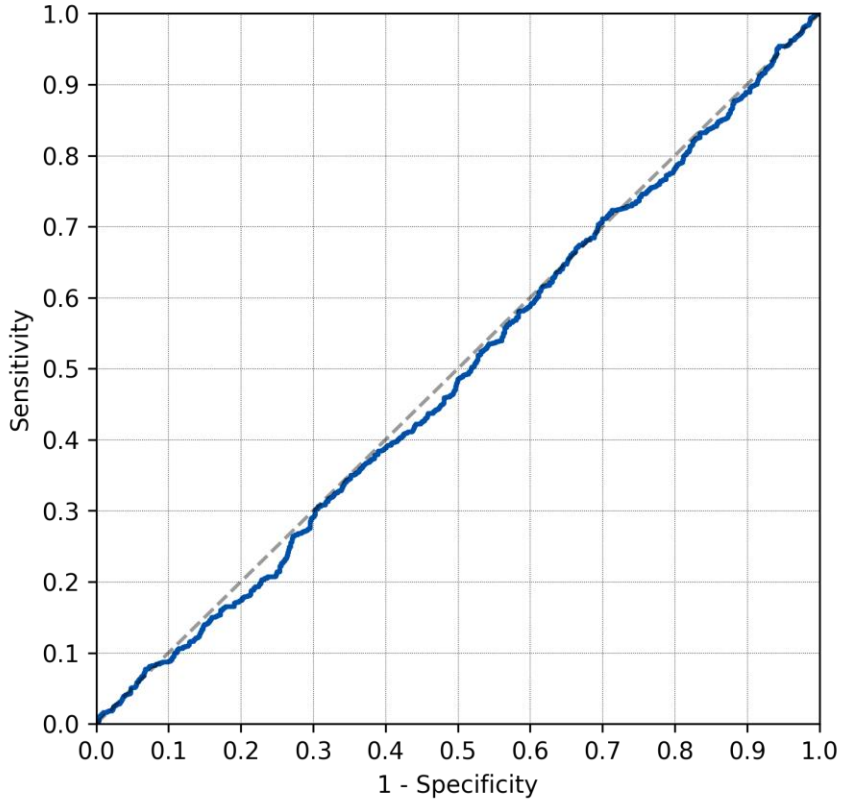
Binormalna krzywa ROC

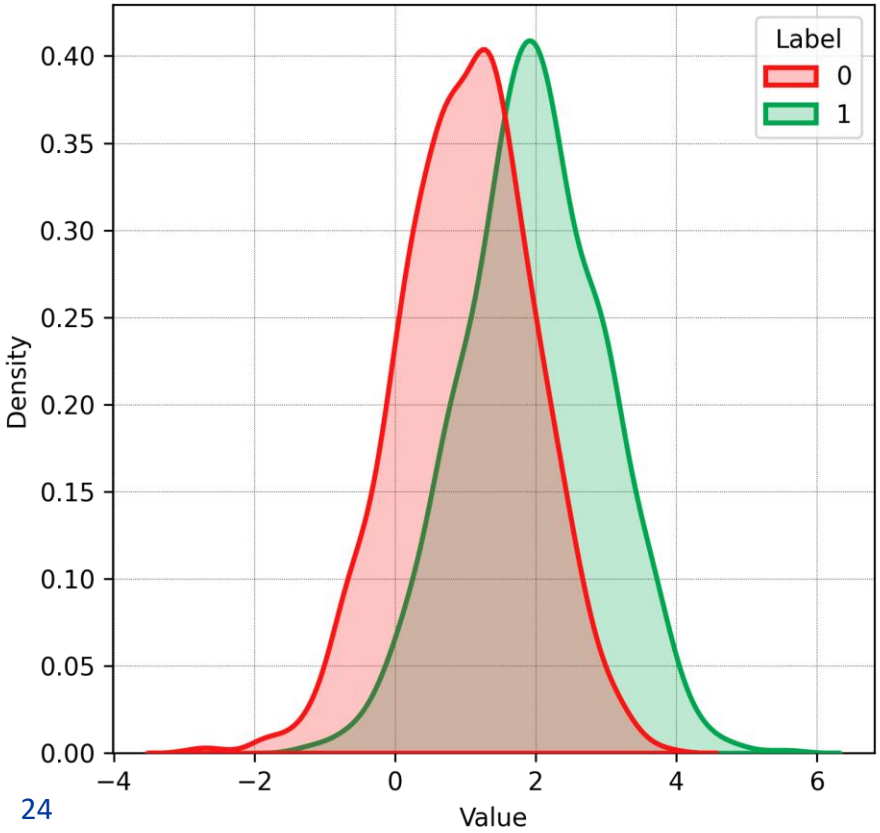
- Jest krzywą gładką. Opiera się na założeniu, że rozkład predyktora (być może po zastosowaniu jakiegoś przekształcenia) w grupach wyznaczonych przez zmienną celu jest normalny.
- W celu wyznaczenia binormalnej krzywej ROC należy wyestymować średnią i odchylenie z próby w grupach. Następnie wygenerować zmienne sztuczne zgodnie z rozkładami normalnymi i powtórzyć kroki pozwalające na wykreślenie krzywej empirycznej.
- Gdy dwa rozkłady normalne nakładają się na siebie, to krzywa jest linią prostą. Gdy rozkłady są idealnie separowalne, to krzywa przechodzi przez punkt (0,1).

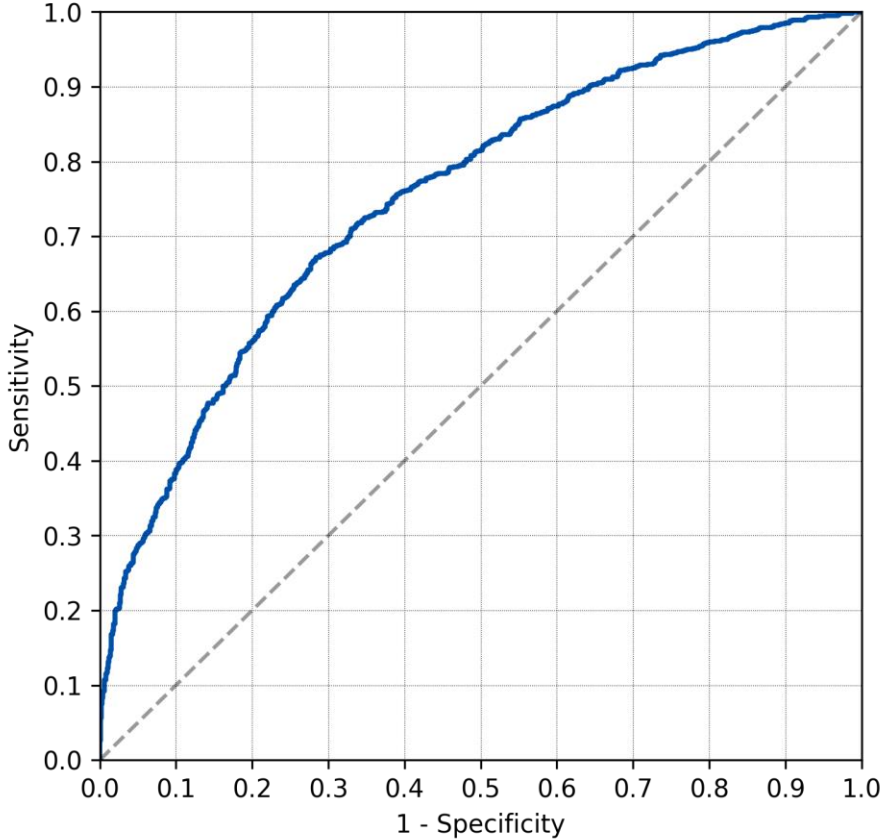


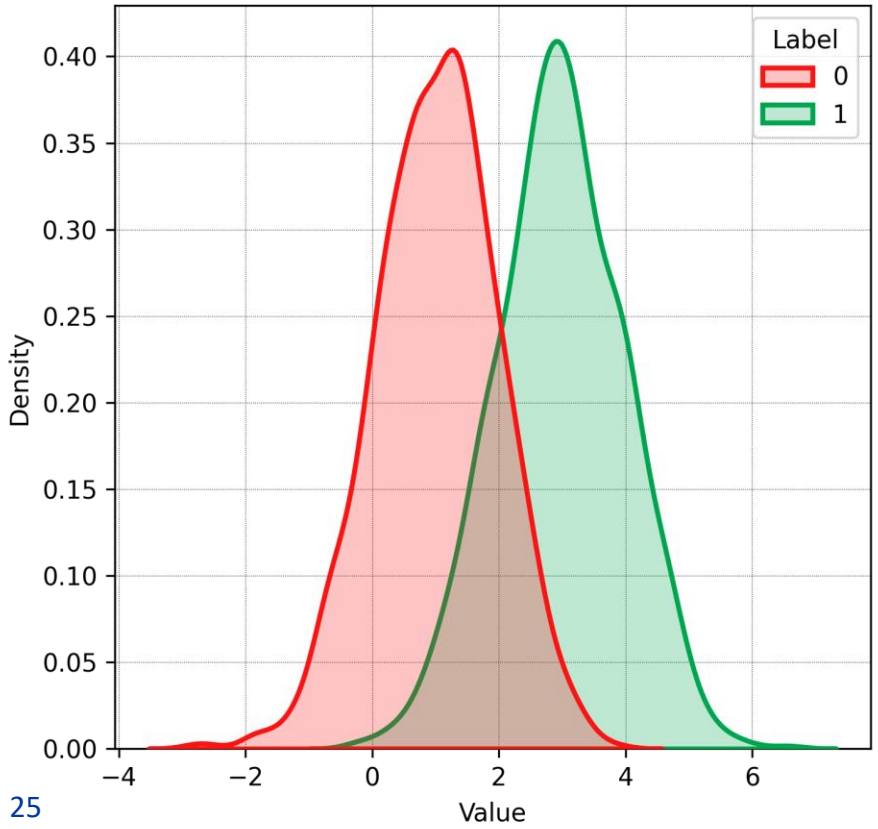
Separowalność rozkładów a kształt krzywej ROC

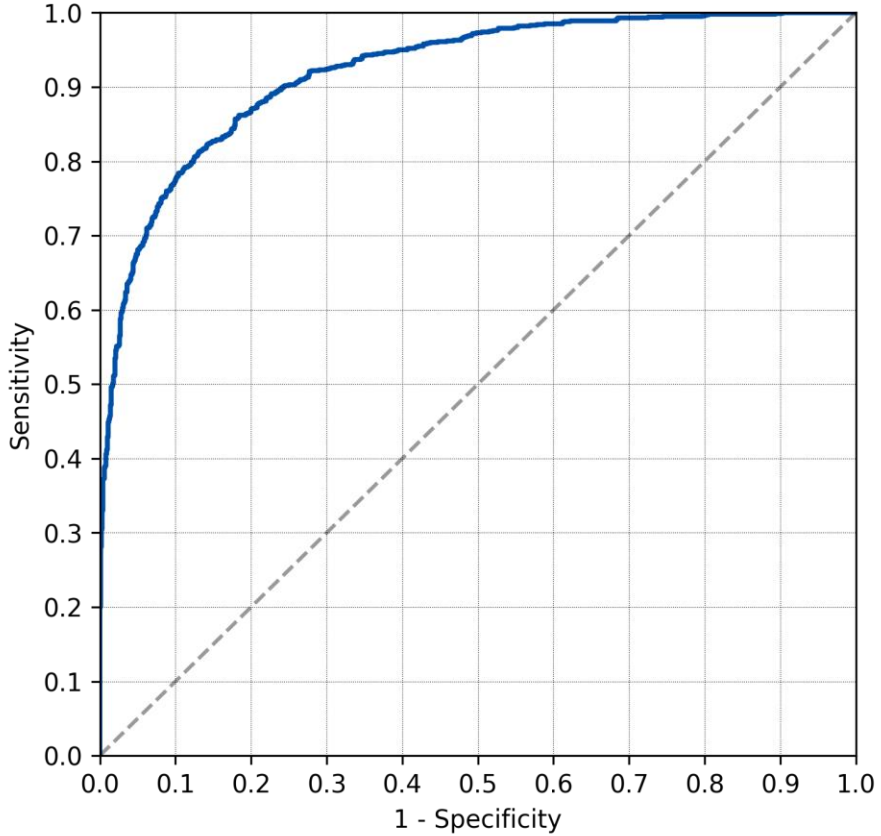


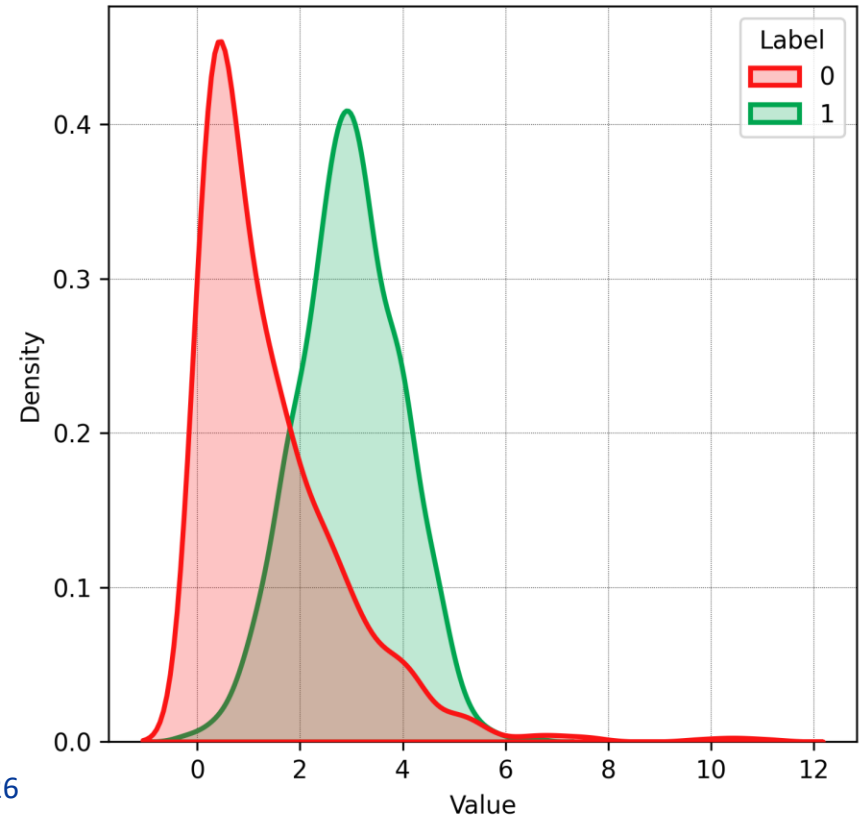


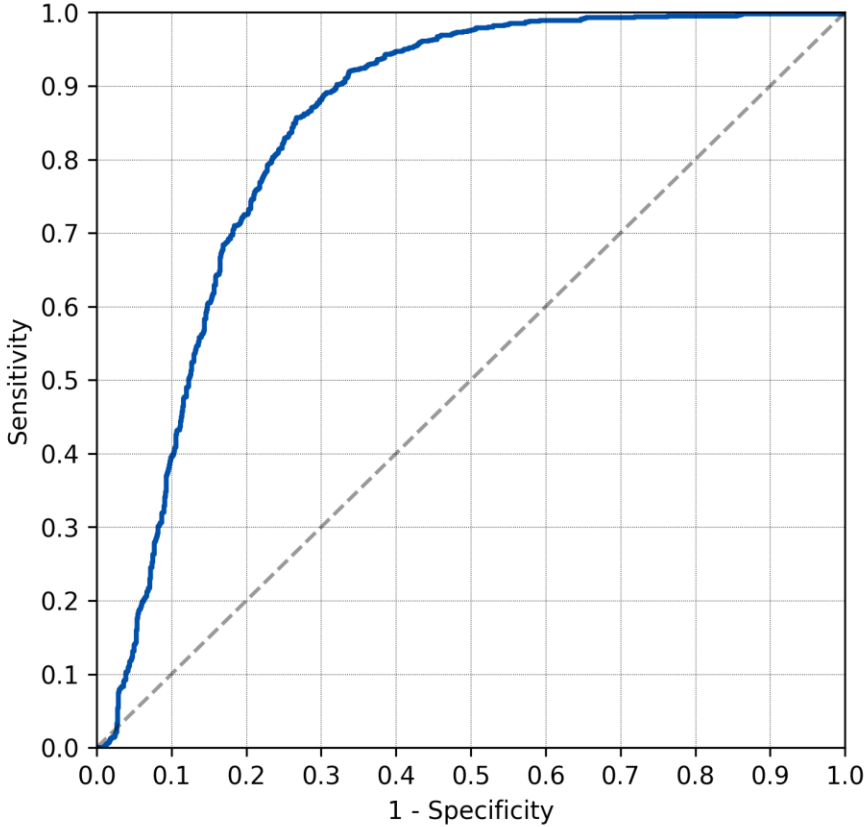

Label
0
1
Density
Value
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
−4
−2
0
2
4
6

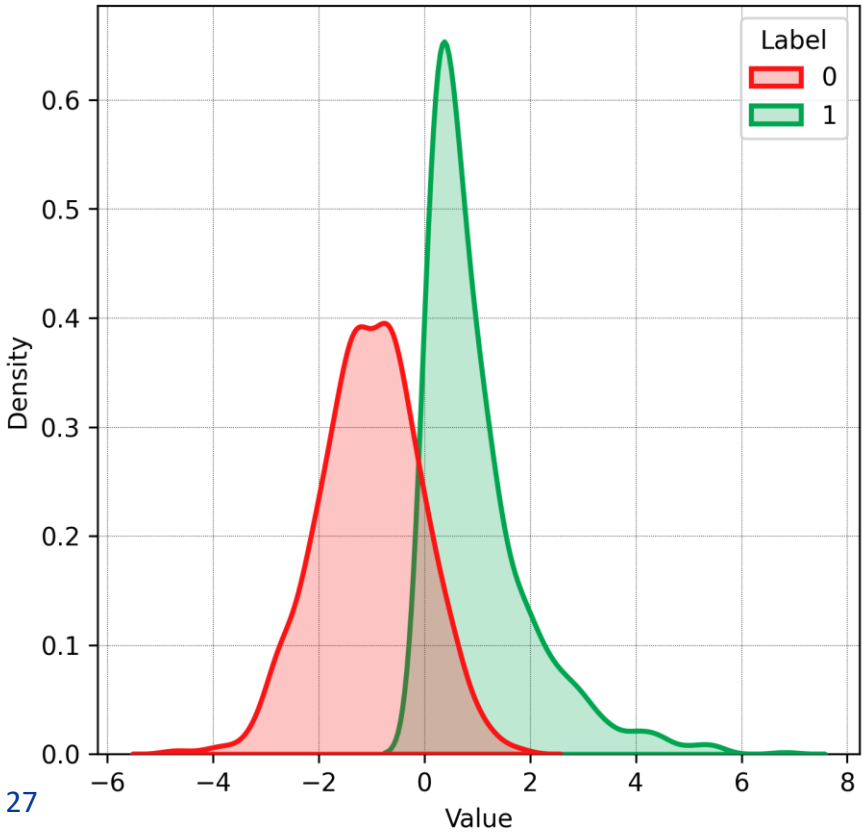

Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

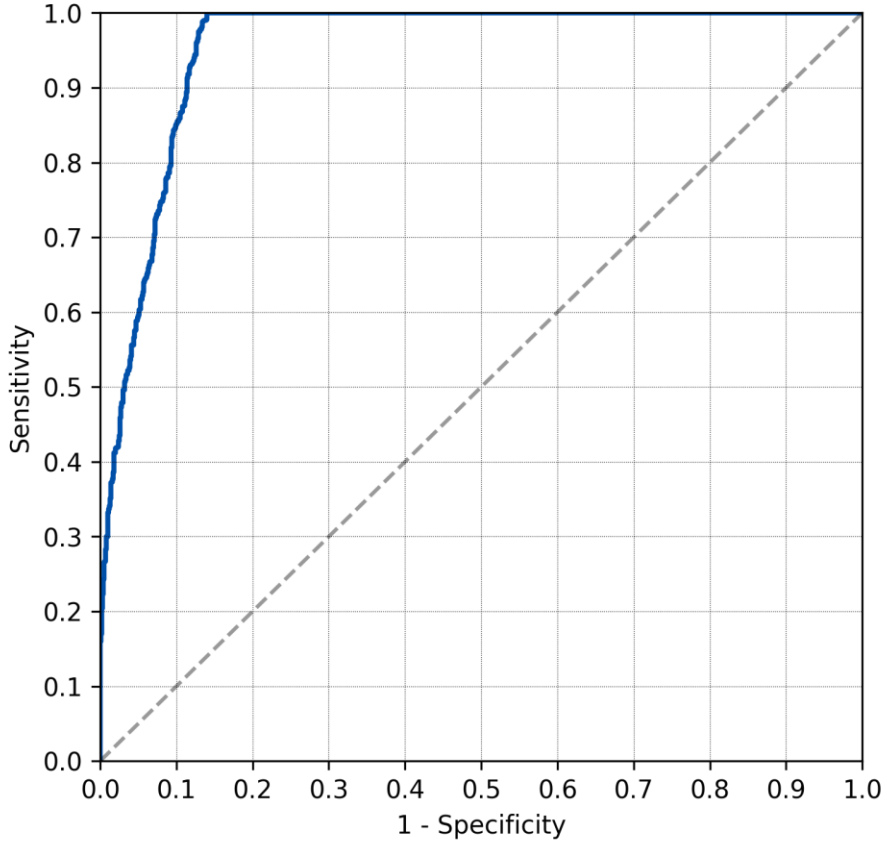

Label
0
1
Density
Value
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
−4
−2
0
2
4
6

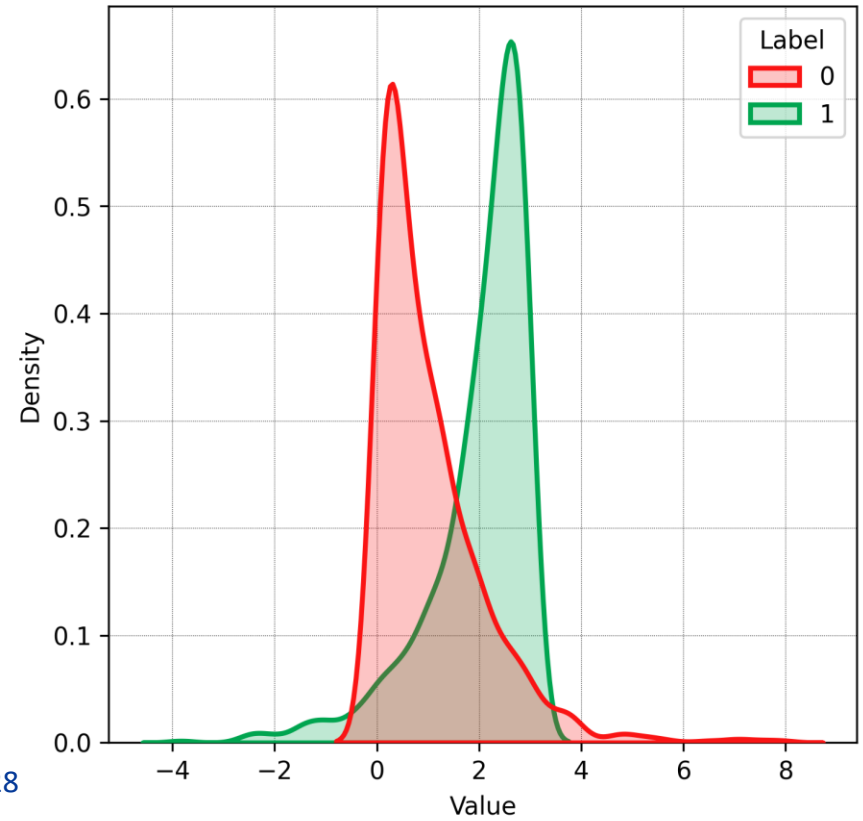

Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

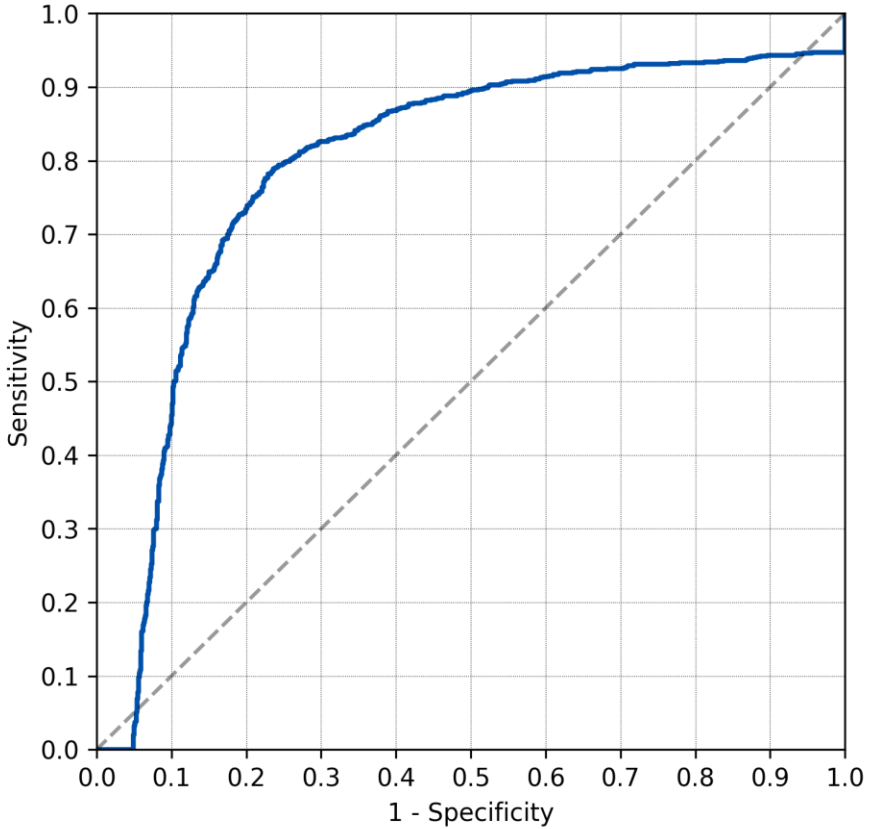

Label
0
1
Density
Value
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
12

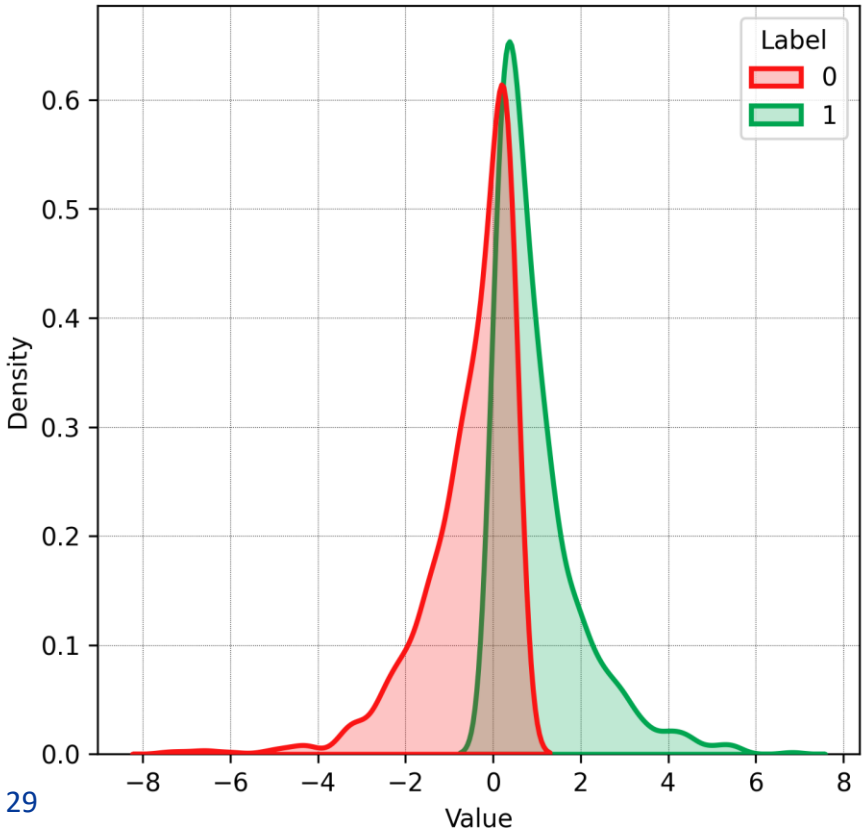

Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

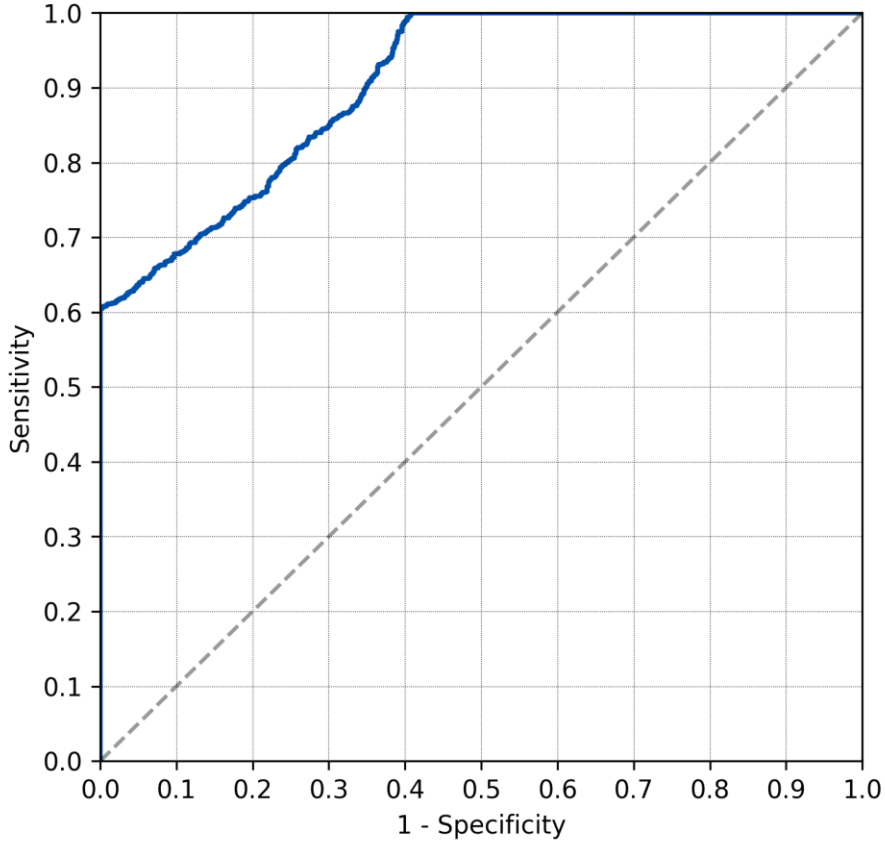

Label
0
1
Density
Value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−6
−4
−2
0
2
4
6
8


Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

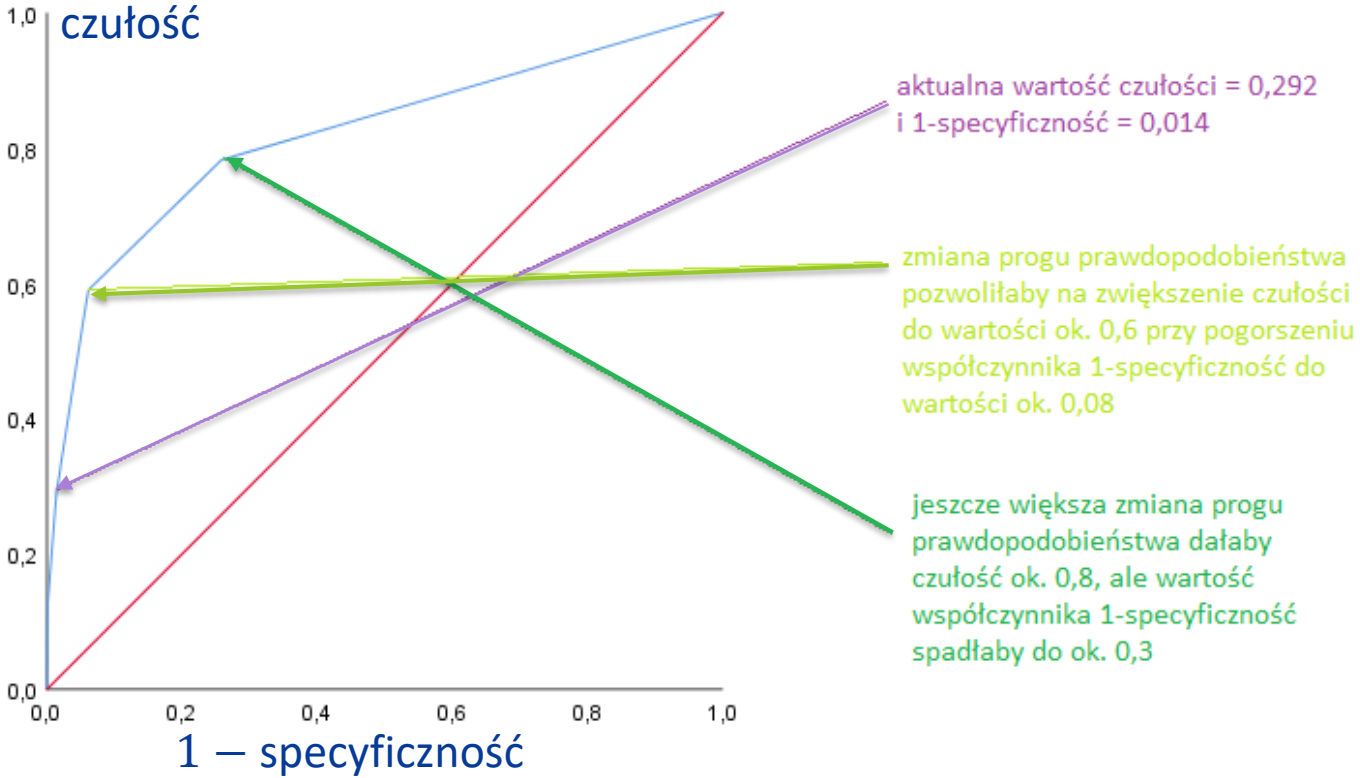

Label
0
1
Density
Value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−4
−2
0
2
4
6
8


Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Label
0
1
Density
Value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−8
−6
−4
−2
0
2
4
6
8


Sensitivity
1 - Specificity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

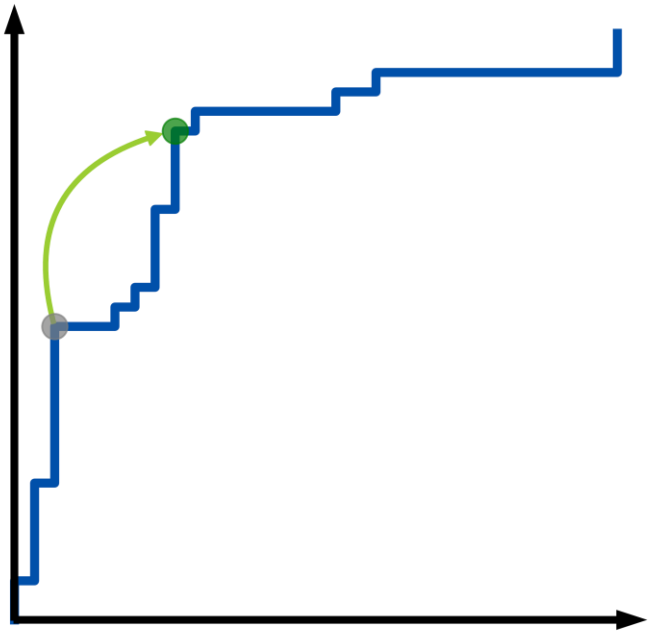
Odnajdując na krzywej punkt uzyskany dla standardowego progu prawdopodobieństwa 0,5, możemy sprawdzić, czy zmiana progu na inny nie pozwoliłaby poprawić jednego wskaźnika (np. czułości) kosztem niewielkiego pogorszenia drugiego.



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

3.

OPTYMALNY PUNKT ODCIĘCIA



Optymalny punkt odcięcia

Przesuwanie się wzdłuż krzywej ROC od lewego dolnego do prawego górnego rogu odpowiada zmniejszaniu progu i powoduje wzrost czułości modelu kosztem pogorszenia jego specyficzności.

Optymalny punkt odcięcia (ang. *optimal cut-off*) jest taką wartością progową prawdopodobieństwa lub predyktora, która zapewnia kompromis pomiędzy czułością i specyficznością modelu.

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Indeks Youdena (ang. *the Youden index*)

(maksymalna metryka Kołmogorowa-Smirnowa)

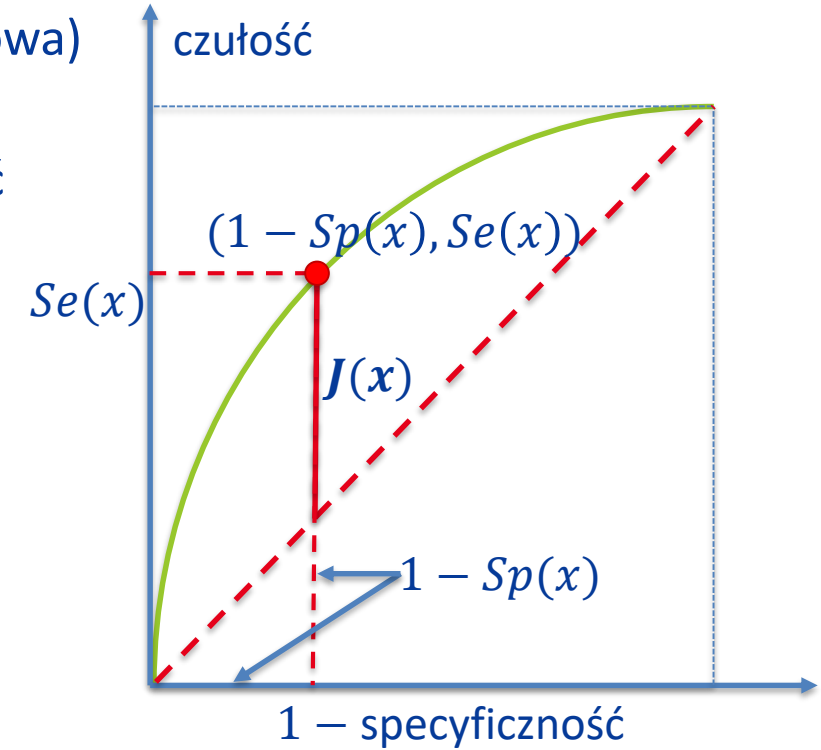
Podany przez Williama J. Youdena (1950).
Wybieramy jako punkt odcięcia taką wartość prawdopodobieństwa / predyktora x, która maksymalizuje funkcję

$$J(x) = Se(x) - \big(1 - Sp(x)\big),$$

Gdzie $Se(x)$ jest czułością przy progu x, a $Sp(x)$ jest specyficznością.

Pozwala to odnaleźć punkt krzywej o największej odległości w pionie od prostej o równaniu

$$\text{czułość} = 1 - \text{specyficzność}.$$



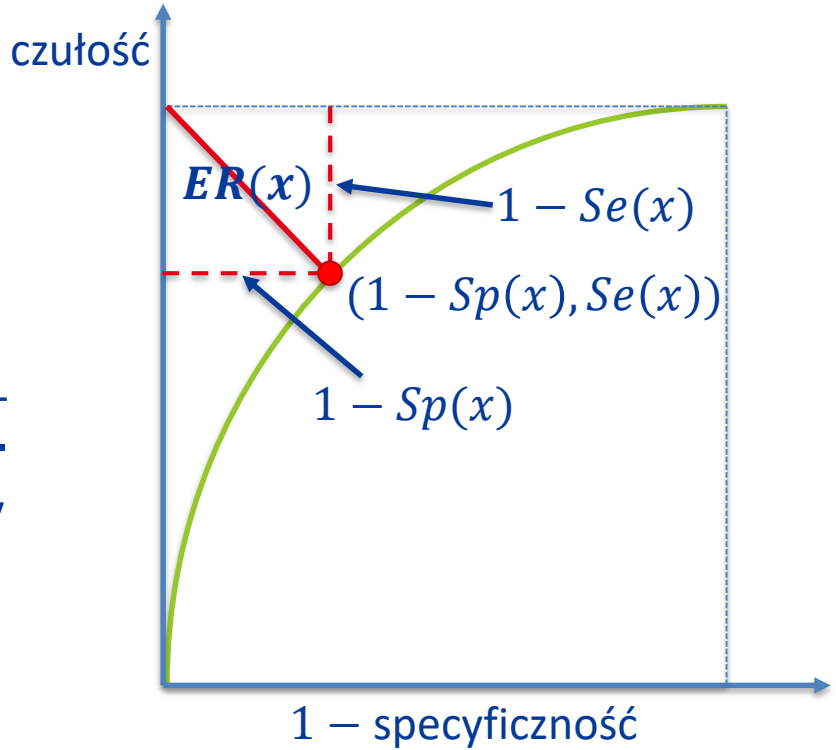
UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Kryterium bycia najbliżej punktu $(0, 1)$

(ang. *the closest to* $(0,1)$ *criteria*)

Wybieramy jako punkt odcięcia taką wartość prawdopodobieństwa / predyktora x, która minimalizuje funkcję $ER(x) = \sqrt{(1 - Se(x))^2 + (1 - Sp(x))^2}$.

Pozwala to odnaleźć punkt krzywej leżący najbliżej punktu $(0,1)$.



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

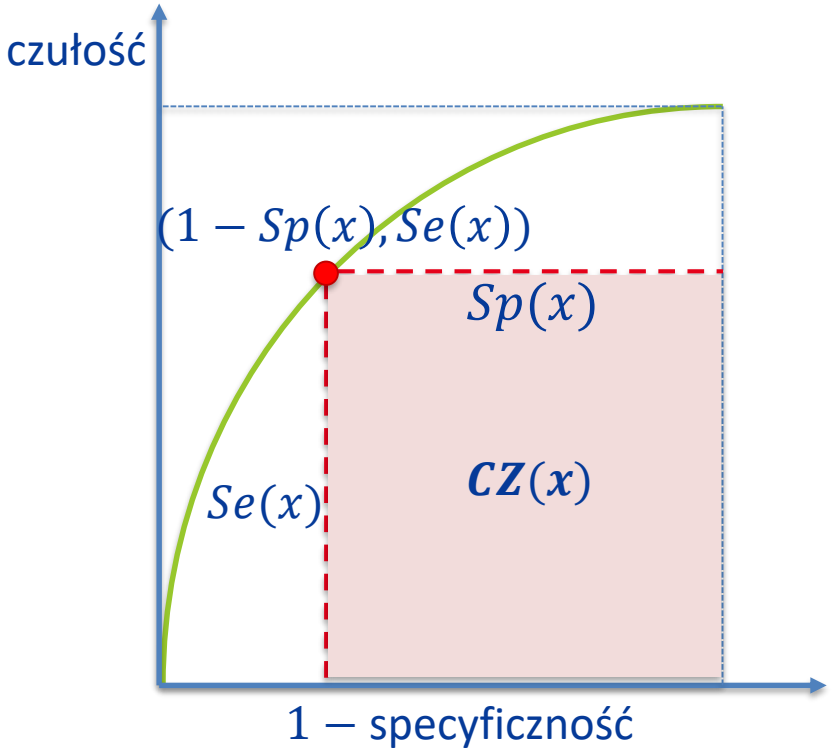
Metoda prawdopodobieństwa zgodności

(ang. *concordance probability method*)

Podana przez X. Liu (2012). Wybieramy jako punkt odcięcia taką wartość prawdopodobieństwa / predyktora x, która maksymalizuje funkcję

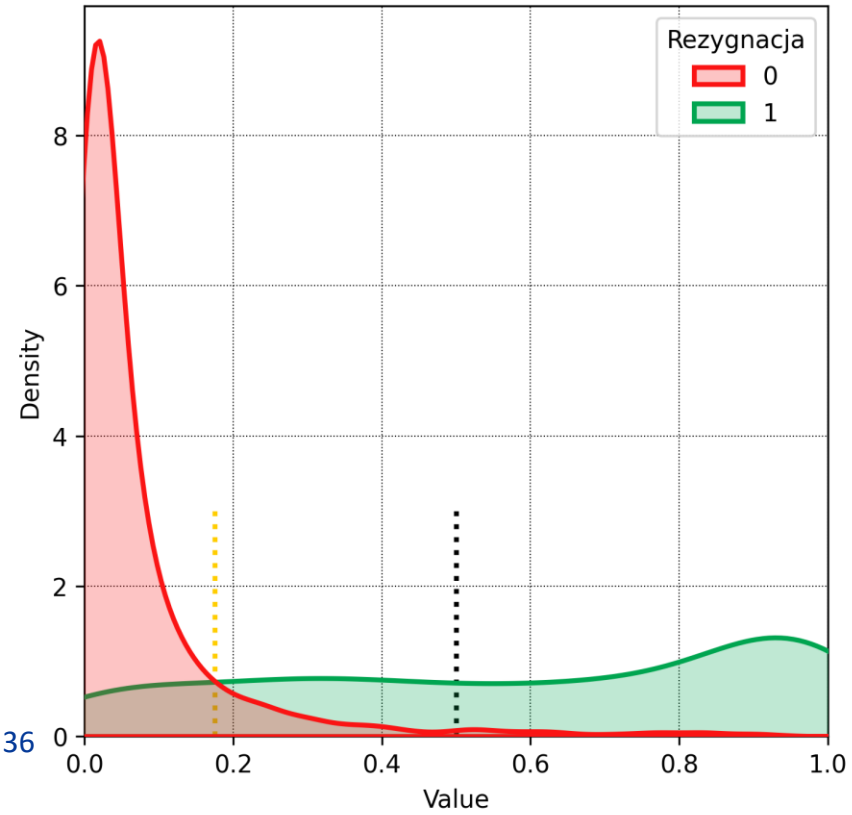
$$CZ(x) = Se(x) \cdot Sp(x).$$

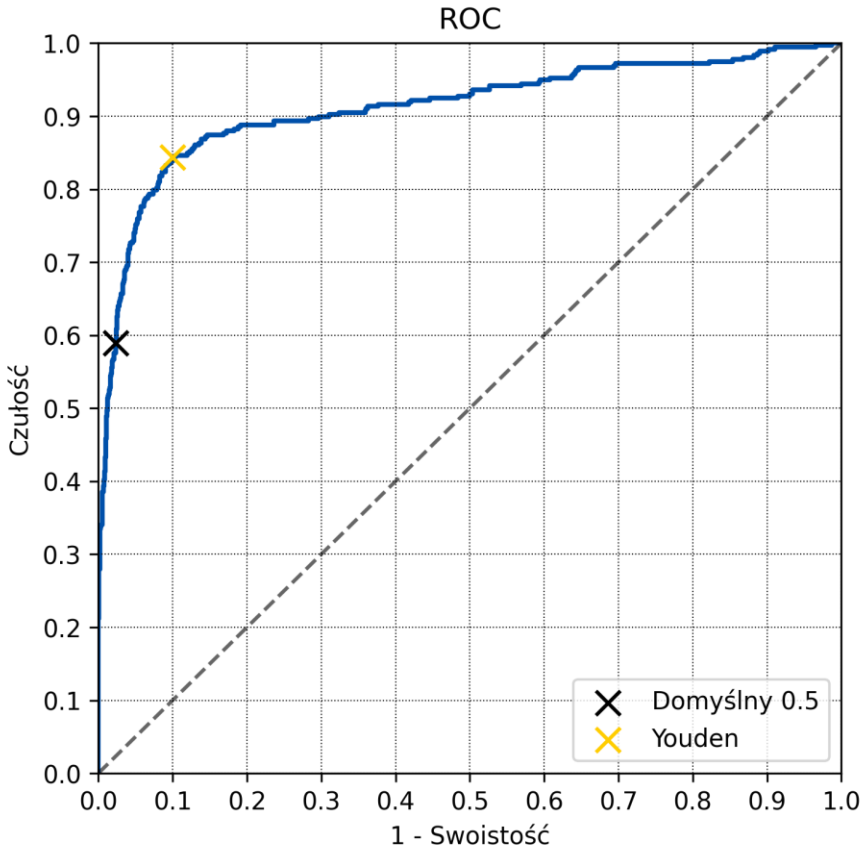
Pozwala to odnaleźć punkt krzywej wyznaczający prostokąt o największym polu.



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Przykład

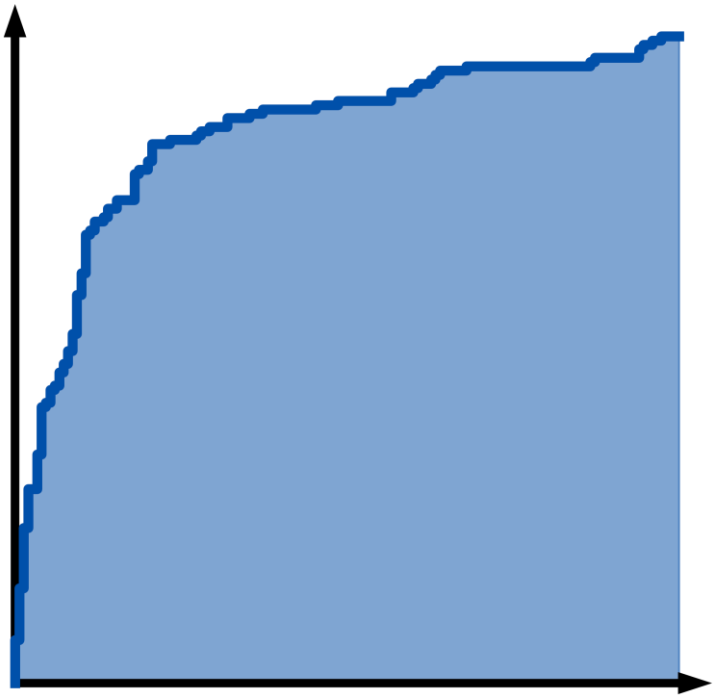




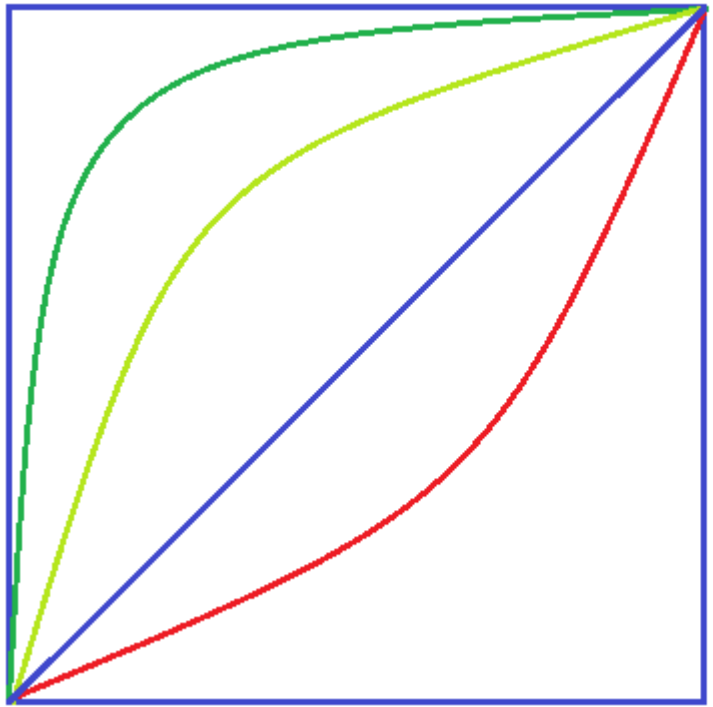
UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

4.

POLE POD KRZYWĄ



Kształt krzywej ROC

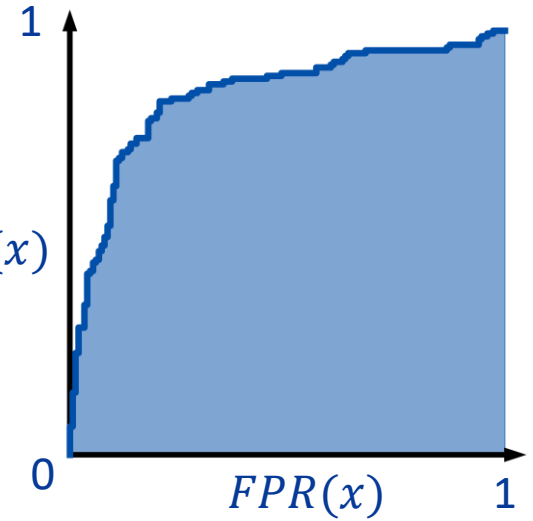


- świetny klasyfikator
- dobry klasyfikator
- linia referencyjna
- błędny klasyfikator (zazwyczaj oznacza, że o przynależności do klasy 1 świadczą coraz to mniejsze, a nie coraz to większe wartości predyktora)

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Pole pod krzywą ROC

(ang. *Area Under Curve*, **AUC**) może być traktowane jako miara sukcesu klasyfikatora i może służyć do porównywania jakości klasyfikatorów.



$$AUC = \int_{-\infty}^{\infty} TPR(x) \cdot FPR'(x)dx$$

Interpretacja AUC:

- estymacja prawdopodobieństwa tego, że wartość X dla losowo wybranej obserwacji z grupy pozytywnych jest większa niż losowo wybranej obserwacji z grupy negatywnych (które to zapisujemy jako $P(X_1 > X_0)$),
- uśredniona czułość po wszystkich wartościach wskaźnika fałszywie pozytywnych.

Wartości AUC a jakość klasyfikatora

Umownie przyjmujemy następującą zależność jakości klasyfikatora od AUC:

- AUC $\in (0{,}6;\ 0{,}7]$ – słaba,
- AUC $\in (0{,}7;\ 0{,}8]$ – akceptowalna (ang. *acceptable*),
- AUC $\in (0{,}8;\ 0{,}9]$ – dobra (ang. *excellent*),
- AUC $\in (0{,}9;\ 1]$ – doskonała (ang. *outstanding*).

Źródło: D. W. Hosmer, S. Lemeshow: *Applied Logistic Regression*, 2nd Ed. Chapter 5, John Wiley and Sons, New York, NY (2000), pp. 160-164.

Nieparametryczna estymacja AUC

- Do estymacji wartości AUC najczęściej używa się metody nieparametrycznej, ze względu na fakt, że nie wymaga ona założenia o normalności rozkładów.
- Metoda opublikowana w 1988 roku przez E. R. DeLong , D. M. DeLong., D. L. Clarke-Pearson.
- Wartość AUC jest szacowana poprzez sumowanie pól wyznaczonych przez trapezy pod empiryczną krzywą ROC.

Określmy dla odpowiednio i-tej obserwacji z grupy pozytywnych (1) i j-tej z negatywnych (0):

$$V(X_{1,i}) = \frac{1}{n_o}\sum_{j=1}^{n_o} f(X_{1,i}, X_{0,j}) \text{ oraz } V(X_{0,j}) = \frac{1}{n_1}\sum_{i=1}^{n_1} f(X_{1,i}, X_{0,j}),$$

gdzie $f(x,y) = \begin{cases} 1 & \text{dla } x > y \\ 0{,}5 & \text{dla } x = y \\ 0 & \text{dla } x < y \end{cases}$ (funkcja Heaviside'a różnicy $x - y$).

Określamy empiryczną wartość AUC jako odsetek par sprzyjających zajściu zdarzenia $X_1 > X_0$:

$$AUC_{emp} = \sum_{i=1}^{n_1} \frac{V(X_{1,i})}{n_1} = \sum_{j=1}^{n_0} \frac{V(X_{0,j})}{n_0} = \frac{1}{n_0 n_1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} f(X_{1,i}, X_{0,j}).$$

Alternatywny sposób:

$$AUC_{emp} = \frac{1}{n_0 n_1} \sum_x \left(n_{TN|X=x} \cdot n_{TP|X>x} + \frac{1}{2} n_{TN|X=x} \cdot n_{TP|X=x} \right)$$

lub

$$AUC_{emp} = \frac{U}{n_0 n_1},$$

gdzie U jest statystyką testową testu U Manna-Whitneya:

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} f(X_{1,i}, X_{0,j}) = R_1 - \frac{1}{2} n_1 (n_1 + 1),$$

a R_1 jest sumą rang wartości zmiennej X w grupie 1.

Uwaga: Gdy małe wartości X świadczą, że $Y = 1$, a duże, że $Y = 0$, to obliczamy powyższą sumę, a następnie odejmujemy od 1.

Szacujemy wariancję

$$Var\left(AUC_{emp}\right) = \frac{1}{n_1} S_{X_1}^2 + \frac{1}{n_0} S_{X_o}^2,$$

gdzie

$$S_{X_k}^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (V(X_{k,i}) - AUC_{emp})^2, \qquad k = 0, 1,$$

(zwykły wzór na wariancję – średni kwadrat odchyleń od średniej).

Stąd **asymptotyczny przedział ufności dla AUC** na poziomie ufności $1 - \alpha$ postaci

$$AUC_{emp} \pm z_{1-\alpha/2} SE\left(AUC_{emp}\right).$$

Nieparametryczny asymptotyczny test dla jednej zmiennej

- Hipoteza zerowa: $AUC = 0{,}5$. Hipoteza alternatywna: $AUC \neq 0{,}5$.
- Przy założeniu hipotezy zerowej AUC_{emp} ma asymptotycznie rozkład normalny.
- Statystyka testowa:

$$\frac{AUC_{emp} - 0{,}5}{SD\left(AUC_{emp}\right)\big|_{AUC=0{,}5}}.$$

Parametryczna estymacja AUC

- Metz (1978), McClish (1989)
- Rzadko stosowana ze względu na założenia.
- Założenie:

$$X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2), \qquad X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2).$$

- Krzywą ROC tworzą punkty postaci

$$\big(FPR(x), TPR(x)\big) = \left(\Phi\left(\frac{\mu_0 - x}{\sigma_o}\right), \Phi\left(\frac{\mu_1 - x}{\sigma_1}\right)\right), \qquad -\infty < x < \infty,$$

gdzie $\Phi(x)$ – dystrybuanta normalnego.

$$AUC = \int_{-\infty}^{\infty} TPR(x) \cdot FPR'(x)dx =$$

$$= \int_{-\infty}^{\infty} \left(\Phi\left(\frac{\mu_1 - x}{\sigma_1}\right) \cdot \varphi\left(\frac{\mu_0 - x}{\sigma_0}\right) \right) dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right),$$

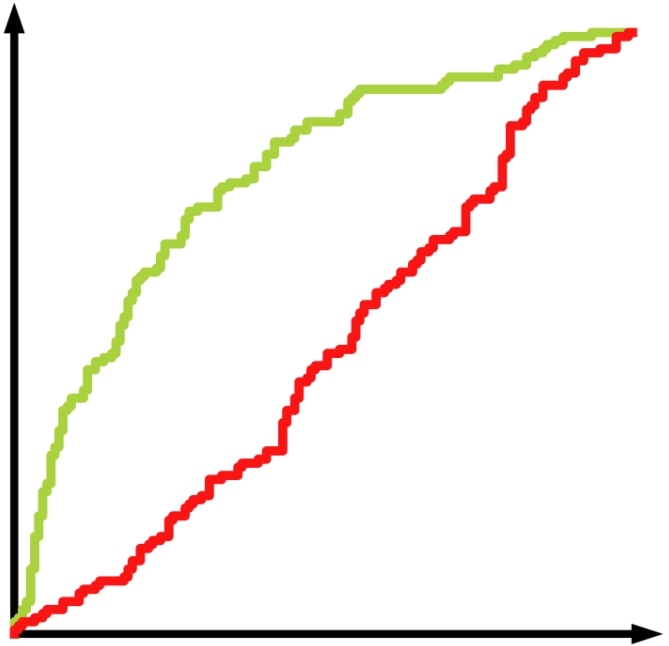
gdzie: $a = \frac{\mu_1 - \mu_0}{\sigma_1} = \frac{\Delta\mu}{\sigma_1}, \quad b = \frac{\sigma_0}{\sigma_1}.$

Dlatego estymatorem AUC jest

$$\widehat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1+\hat{b}^2}}\right).$$

5.

TESTY STATYSTYCZNE DLA DWÓCH KRZYWYCH ROC



Test dla prób niezależnych

- Porównujemy jakość predykcji w oparciu o tą samą zmienną/model w dwóch grupach obserwacji.
- Hipoteza zerowa: $AUC_1 = AUC_2$. Hipoteza alternatywna: $AUC_1 \neq AUC_2$.
- Statystyka testowa:

$$z = \frac{AUC_1 - AUC_2}{\sqrt{\left(SE(AUC_1)\right)^2 + \left(SE(AUC_2)\right)^2}}$$

- Przy założeniu hipotezy zerowej z ma asymptotycznie rozkład normalny.

Test DeLongów dla prób zależnych (sparowanych)

- Porównujemy jakość predykcji w oparciu o różne zmienne/modele .
- Hipoteza zerowa: $AUC_1 = AUC_2$. Hipoteza alternatywna: $AUC_1 \neq AUC_2$.
- Statystyka testowa:

$$z = \frac{AUC_1 - AUC_2}{Var(AUC_1 - AUC_2)},$$

gdzie

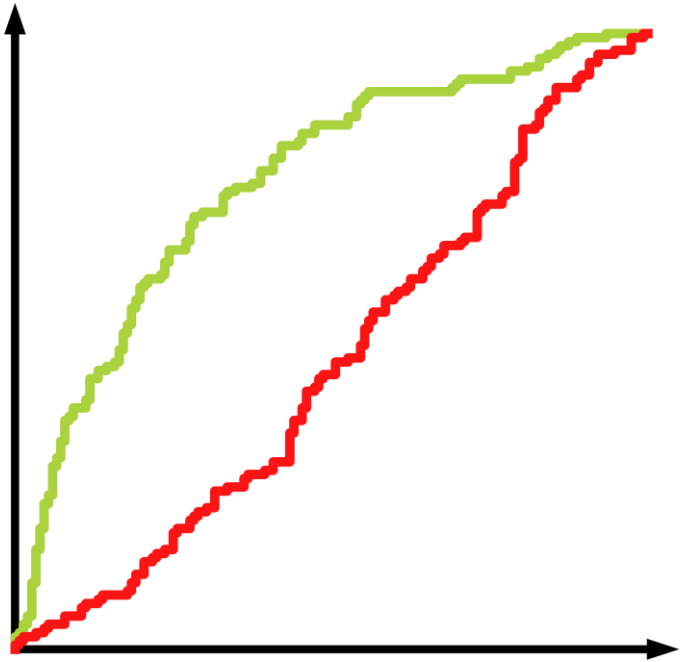
$$Var(AUC_1 - AUC_2) = Var(AUC_1) + Var(AUC_2) - 2Cov(AUC_1, AUC_2),$$

$$Cov(AUC_1, AUC_2) = \frac{S_{X_{1,1}X_{2,1}}}{n_1} + \frac{S_{X_{1,0}X_{2,0}}}{n_0},$$

$$S_{X_{1,k}X_{2,k}} = \frac{1}{n_k - 1}\sum_{j=1}^{n_k}\left(V\left(X_{1,k,j}\right) - AUC_1\right)\left(V\left(T_{2,k,j}\right) - AUC_2\right),\ k = 0,1.$$

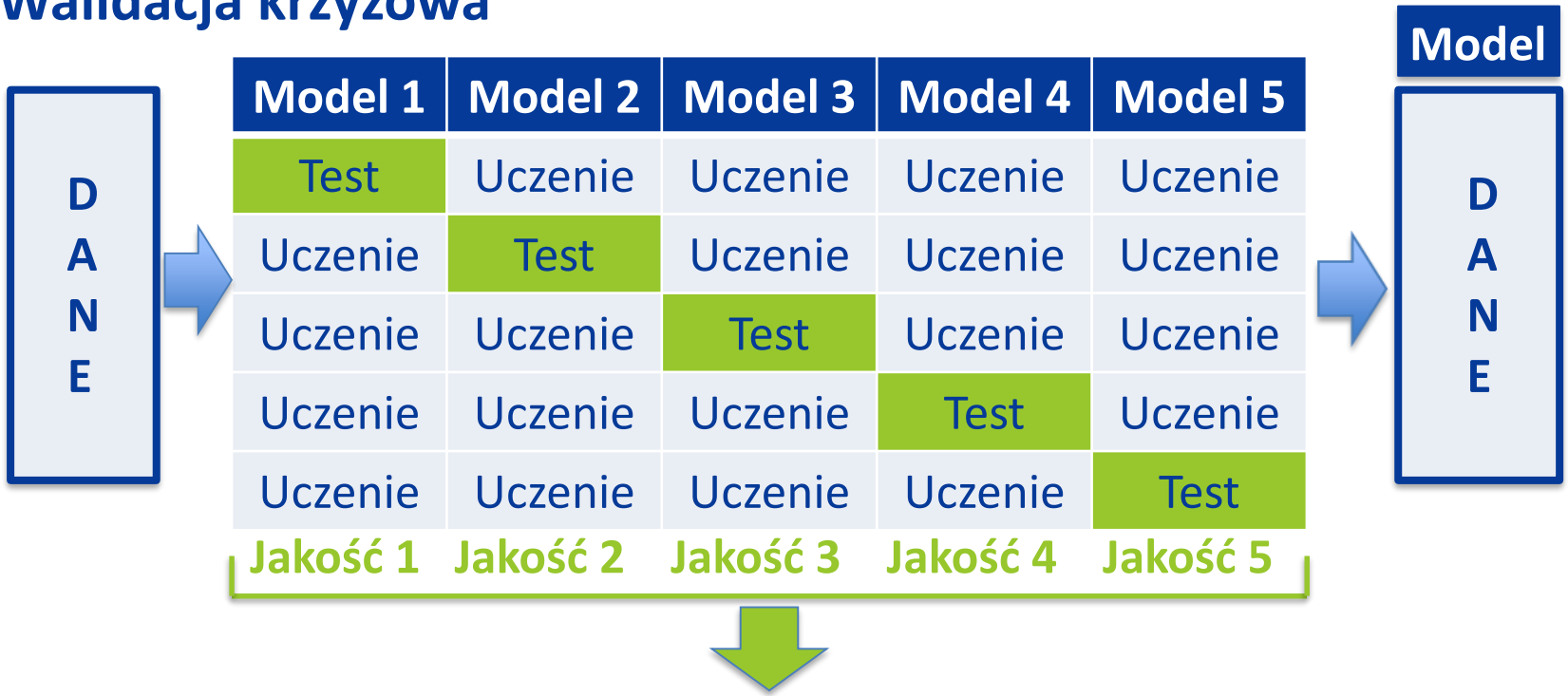
- Przy założeniu hipotezy zerowej z ma asymptotycznie rozkład normalny.

6. TROCHĘ PRAKTYKI



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Walidacja krzyżowa



Jakość = średnia(Jakość 1, …, Jakość 5)

Strategie obliczania AUC w przypadku walidacji krzyżowej:

1. **Uśrednianie** (ang. *averaging*) – wyznaczamy krzywą ROC i obliczamy AUC na każdym ze zbiorów testowych z osobna, a następnie uśredniamy wynik.
2. **Łączenie** (ang. *pooling*) – zbieramy wyniki działania klasyfikatora na każdym ze zbiorów testowych i łączymy je w jeden duży zbiór. Dla tego zbioru rysujemy krzywą ROC i obliczamy AUC. Ta metoda jest łatwiejsza w implementacji i jest jedyną możliwą do zastosowania w przypadku N-krotnej walidacji krzyżowej (*leave-one-out*).

Uśrednianie

DANE

Model 1	Model 2	Model 3	Model 4	Model 5
Test	Uczenie	Uczenie	Uczenie	Uczenie
Uczenie	Test	Uczenie	Uczenie	Uczenie
Uczenie	Uczenie	Test	Uczenie	Uczenie
Uczenie	Uczenie	Uczenie	Test	Uczenie
Uczenie	Uczenie	Uczenie	Uczenie	Test
AUC 1	AUC 2	AUC 3	AUC 4	AUC 5

AUC = średnia(AUC 1, …, AUC 5)

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Łączenie

Model 1	Model 2	Model 3	Model 4	Model 5	P-stwo
Test	Uczenie	Uczenie	Uczenie	Uczenie	Test
Uczenie	Test	Uczenie	Uczenie	Uczenie	Test
Uczenie	Uczenie	Test	Uczenie	Uczenie	Test
Uczenie	Uczenie	Uczenie	Test	Uczenie	Test
Uczenie	Uczenie	Uczenie	Uczenie	Test	Test

AUC

Strategie obliczania AUC w przypadku klasyfikacji niebinarnej

1. **Odniesienie do klasy** (ang. *class reference*) – wyznaczamy krzywą ROC dla każdej z klas, traktując ją jako wartość pozytywną, a pozostałe klasy łącznie jako jedną klasę negatywną.
2. **Podejście Provosta i Domingosa** (2001)

$$AUC_{total} = \sum_{c_i \in C} AUC(c_i) p(c_i),$$

gdzie $AUC(c_i)$ jest określone jak w 1., a $p(c_i)$ jest częstością klasy c_i.

3. **Podejście Handa i Hilla** (2001)

$$AUC_{total} = \frac{2}{|C|(|C|-1)} \sum_{\{c_i, c_j\} \in C} AUC(c_i, c_j),$$

gdzie $AUC(c_i, c_j)$ odnosi się do klasyfikacji binarnej do klas c_i, c_j.

Python

- *sklearn.metrics.roc_curve* – oblicza współrzędne krzywej ROC (tylko dla klasyfikacji binarnej). Zwraca 3 tablice: czułości, wskaźniki fałszywie pozytywne i wartości odcięcia.
- *sklearn.metrics.roc_auc_score* – oblicza wartość AUC nawet dla klasyfikacji wieloklasowej i wieloetykietowej (z możliwością wybrania typu średniej: *micro*, *macro*, *weighted*).
- ~~*sklearn.metrics.plot_roc_curve*~~
  sklearn.metrics.RocCurveDisplay – wizualizacja krzywej ROC, w oparciu o wartości przewidywane (metoda *from_predictions*) lub poprzez przekazanie modelu (metoda *from_estimator*) bez zwracania wartości tpr, fpr, thr.

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Optymalny punkt odcięcia

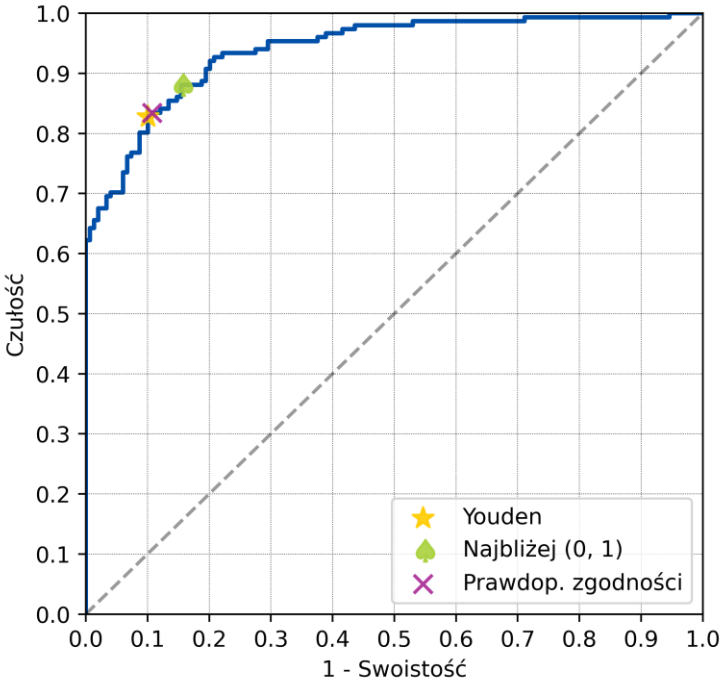
```
def best_threshold(tpr, fpr, thr, method='youden'):

    if method == 'youden':
        pkt = np.argmax(tpr-fpr)
        return tpr[pkt], fpr[pkt], thr[pkt]

    elif method == 'closest_01':
        pkt = np.argmin(np.sqrt((1-tpr)**2 + fpr**2))
        return tpr[pkt], fpr[pkt], thr[pkt]

    elif method == 'con_probab':
        pkt = np.argmax(tpr * (1-fpr))
        return tpr[pkt], fpr[pkt], thr[pkt]

    .
    .
    .
```



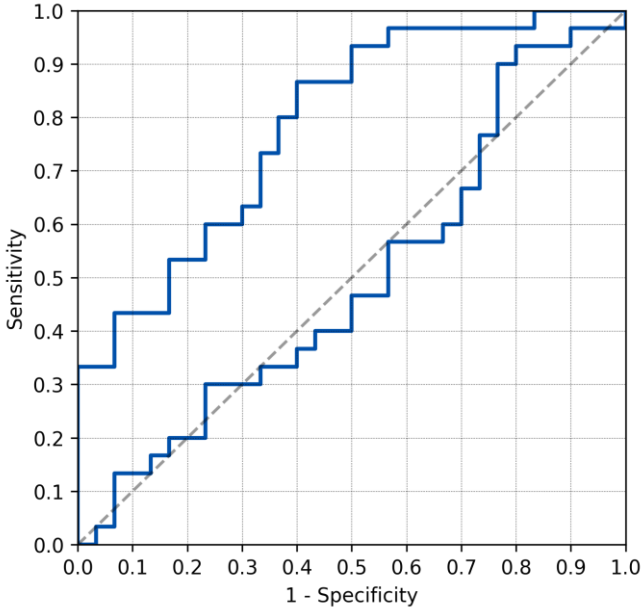
UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Test DeLongów dla prób zależnych (sparowanych)

- Test: https://github.com/yandexdataschool/roc_comparison/blob/master/compare_auc_delong_xu.py

```
p_val = 10**compare_auc_delong_xu\
    .delong_roc_test(y, X[:,0], X[:,1])[0][0]
```

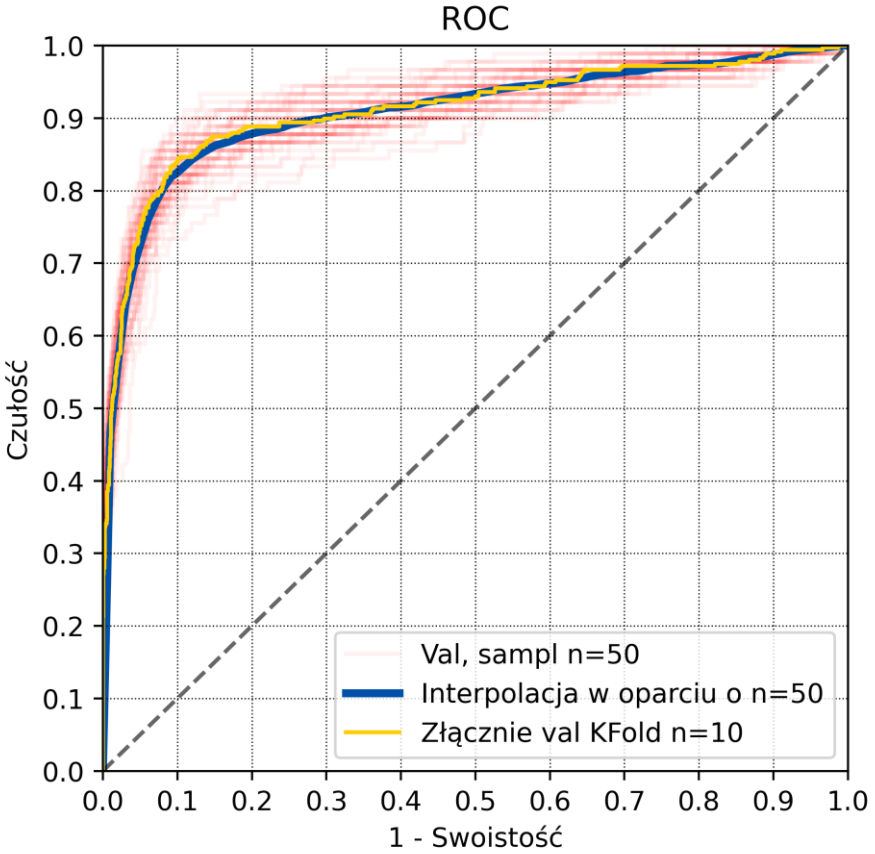
- Można całkiem łatwo samodzielnie zaimplementować.



UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Interpolacja krzywej

1. Przeprowadzamy wielokrotną walidację (najlepiej przez wielokrotne samplowanie).
2. Dla każdej iteracji wyznaczamy krzywą.
3. W oparciu o zestaw krzywych interpolujemy krzywą uśrednioną (w pythonie można użyć *scipy.interp* – w wersji niższej niż 2.0.0).



Środowisko R

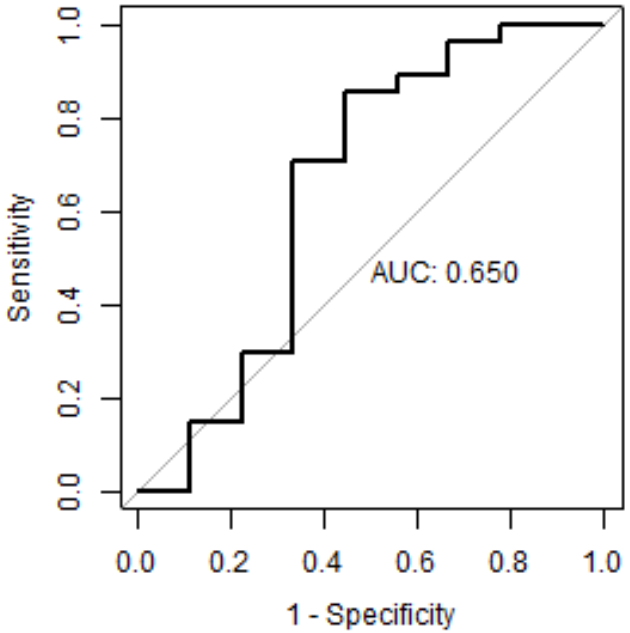
Najpopularniejsze biblioteki do generowania krzywych ROC:

- pROC,
- plotROC,
- ROCR.

Do wyznaczania punktów odcięcia można wykorzystywać dodatkową biliotekę *OptimalCutpoints.*

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Biblioteka pROC



(opracowanie J. Wojtasik)

Możliwości:

- Proste generowanie wykresów i ich personalizacja:

```
plot(roc, print.AUC = T,
legacy.axes = T)
```

- Wyznaczanie wartości współczynnika AUC:

```
auc(roc)
```

https://rdocumentation.org/packages/pROC/versions/1.18.0

Biblioteka pROC

- Wyznaczanie punktów odcięcia

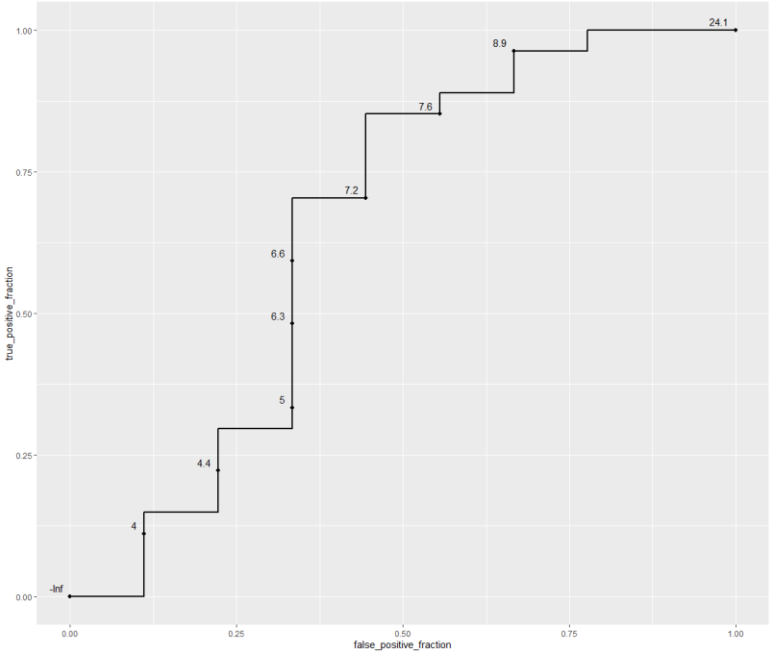
```
coords(roc, "best", best.method="youden")
```

- Test DeLongów dla prób zależnych (sparowanych)

```
roc.test(roc1, roc2)
```

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

Inne biblioteki



(opracowanie J. Wojtasik)

plotROC:

- Generowane z wykorzystaniem biblioteki *ggplot2*,
- Możliwość wyznaczenia AUC

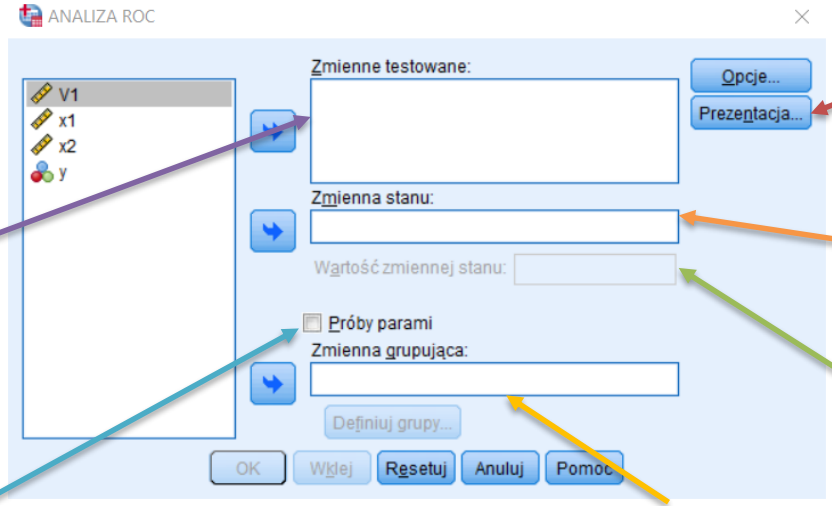
```
calc_auc(plot)
```

http://pbiecek.github.io/Przewodnik/Predykcja/ROC.html

UNIWERSYTET MIKOŁAJA KOPERNIKA W TORUNIU

PS IMAGO PRO

Analiza → Klasyfikacja → Analiza ROC



W „Prezentacja…” można zaznaczyć opcje:
- ogólnej jakości modelu,
- wyznaczenia wszystkich współrzędnych krzywej*,
- obliczenia błędu i przedziałów ufności,
- obliczenia optymalnego punktu odcięcia (metodą Youdna).

Tutaj umieszczamy predyktory

Tutaj umieszczamy zmienną stanu (celu)

Wpisujemy wartość kategorii, którą uznajemy za główny cel

Zaznaczenie powoduje wykonanie testu dla prób sparowanych, a nie dla prób niezależnych

Opcjonalne. Umieszczamy tu zmienną grupującą w przypadku chęci wykonania testu dla prób niezależnych

* Program oblicza współrzędne, jednak wartości odcięcia są średnią arytmetyczną dwóch najbliższych punktów

Bibliografia

- Tom Fawcett: *An introduction to ROC analysis*. Pattern Recognition Letters 27 (2006), pp. 861–874.
- IBM SPSS Statistics Algorithms
- Rachel Draelos: *Comparing AUCs of Machine Learning Models with DeLong's Test*. https://glassboxmedicine.com/2020/02/04/comparing-aucs-of-machine-learning-models-with-delongs-test/ (dostęp z dnia 19.05.2022)
- E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson: *Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.* Biometrics Vol. 44, No. 3 (Sep., 1988), pp. 837-845
- D. Katzman McClish *Analyzing a Portion of the ROC Curve*. Medical Decision Making, Vol. 9 (3), 1989, 190–195.