

Ocena metod klasyfikacji (Evaluation methods)

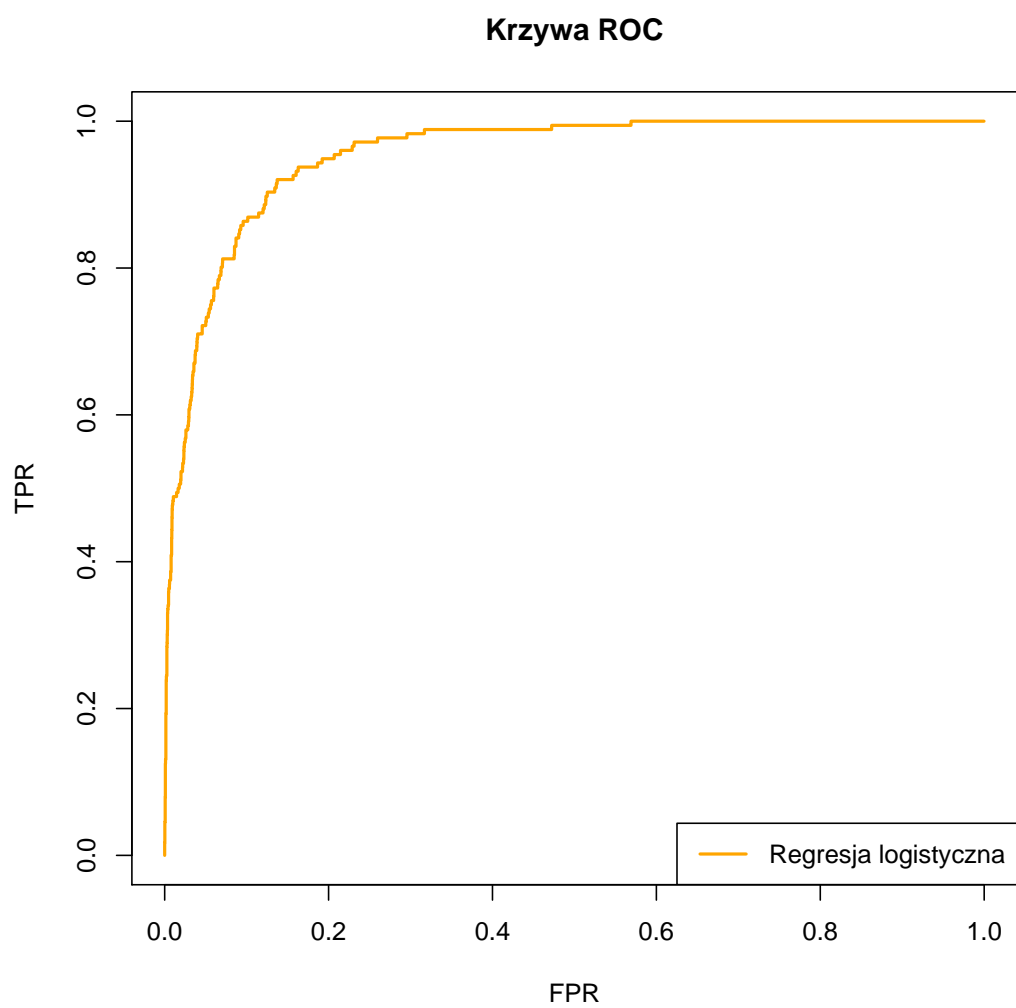
Zadania:

1. Podstawowe miary oceny klasyfikatorów.

- Wczytaj dane `default` z pakietu `ISLR`.
- Podziel dane losowo na zbiór treningowy (50%) i testowy (50%).
- Dopasuj 2 modele klasyfikacji na zbiorze treningowym: regresja logistyczna (funkcja `glm`) i drzewo klasyfikacyjne (funkcja `rpart`).
- Oblicz następujące wskaźniki na zbiorze testowym dla rozważanych modeli:
 - Dokładność (Accuracy) = $\frac{\#\{\hat{y}=y\}}{n}$,
 - Czułość (Recall, True Positive Rate) = $\frac{\#\{\hat{y}=+,y=+\}}{\#\{y=+\}}$,
 - Precyzja (Precision) = $\frac{\#\{\hat{y}=+,y=+\}}{\#\{\hat{y}=+\}}$,
gdzie n to licznosc zbioru testowego, y prawdziwa wartosc zmiennej odpowiedzi, \hat{y} przewidywana wartosc zmiennej odpowiedzi. Przyjmujemy że $\hat{y} = +$ jeżeli $\hat{P}(y = +|x) > 0.5$ (\hat{P} to prawdopodobienstwo aposteriori oszacowane na podstawie modelu).
- Oblicz na zbiorze testowym następujący wskaźnik dla rozważanych modeli:
 - Precyzja @ $k\%$ (Precision @ $k\%$) = $\frac{\#\{\hat{y}=+,y=+\}}{\#\{\hat{y}=+\}}$,
gdzie $\hat{y} = +$ dla $k\%$ największych wartosci $\hat{P}(y = +|x)$.

2. Własna funkcja do rysowania krzywej ROC.

- Krzywa ROC powstaje w następujący sposób.
 - Na osi OY oznaczamy True Positive Rate (Recall) $TPR = \frac{\#\{\hat{y}=+,y=+\}}{\#\{y=+\}}$.
 - Na osi OX oznaczamy False Positive Rate $FPR = \frac{\#\{\hat{y}=+,y=-\}}{\#\{y=-\}}$.
 - Przyjmujemy regule klasyfikacyjną $\hat{y} = +$, jeżeli $\hat{P}(y = +|x) > t$, gdzie $t \in (0, 1)$ jest parametrem.
 - Krzywa powstaje przez utworzenie punktów (FPR, TPR) dla różnych wartosci parametru t .
- Napisz własną funkcję do rysowania krzywej ROC. Funkcja powinna pobierać dwa argumenty: wektor zmiennych odpowiedzi dla danych testowych oraz wektor prawdopodobienstw aposteriori dla danych testowych, wyliczony na podstawie dopasowanego modelu.
- Wywołaj funkcję na danych `Default`, dla klasyfikatorów rozważanych w zadaniu 1. Dane treningowe używamy do zbudowania modelu, zaś dane testowe do narysowania krzywej. Rysunek 1 pokazuje krzywą ROC dla regresji logistycznej.



Rysunek 1: Krzywa ROC dla regresji logistycznej, dla danych `Default`.