

Predykcja konformalna - Team Green

Jakub Gontarz, Antoni Szustakowski, Roman Uhera

Warsztaty badawcze PRiMO 2024

czerwiec 2024

Celem poniższego projektu było zaimplementowanie metod predykcji konformalnej do trzech zadań:

- klasyfikacji binarnej na podstawie scorów modelu oraz rzeczywistych klas,
- regresji scoru dla każdej predykcji z modelu,
- regresji scoru dla każdej obserwacji na podstawie wszystkich zmiennych predykcyjnych.

We wszystkich zadaniach *score* oznaczał prawdopodobieństwo klasy 1.

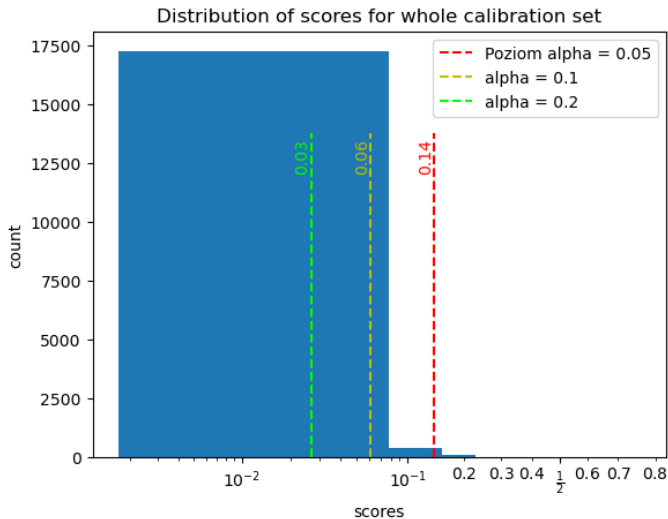
Zadanie 1

W pierwszej części zadania 1 w podstawowej wersji konformalnej predykcji dopasowano kwantyl rozkładu scorów dla wszystkich klientów tak, aby:

$$P\left(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}})\right) \geq 1 - \alpha.$$

Dla klasy 0 *scorem niepewności* jest wynik z modelu. Dla klasy 1 *scorem niepewności* jest 1-score z modelu.

Podstawowa klasyfikacja konformalna



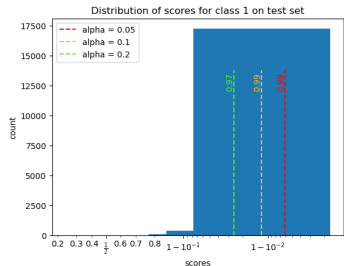
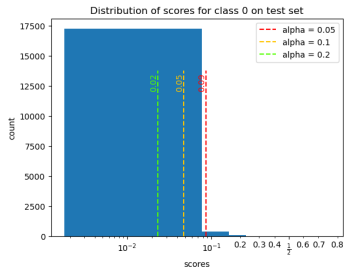
Zazwyczaj nie chcemy osiągnąć zadanego pokrycia jedynie *średnio* po wszystkich zbiorach testowych. W związku z tym dążono do osiągnięcia pokrycia warunkowo pod warunkiem klasy, do której przynależy obserwacja ze zbioru testowego:

$$\mathbb{P}(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}}) \mid Y_{\text{test}} = y) \geq 1 - \alpha,$$

dla $y \in \{0, 1\}$. W tym celu dopasowano kwantyle scorów w zbiorze kalibracyjnym dla każdej klasy osobno.

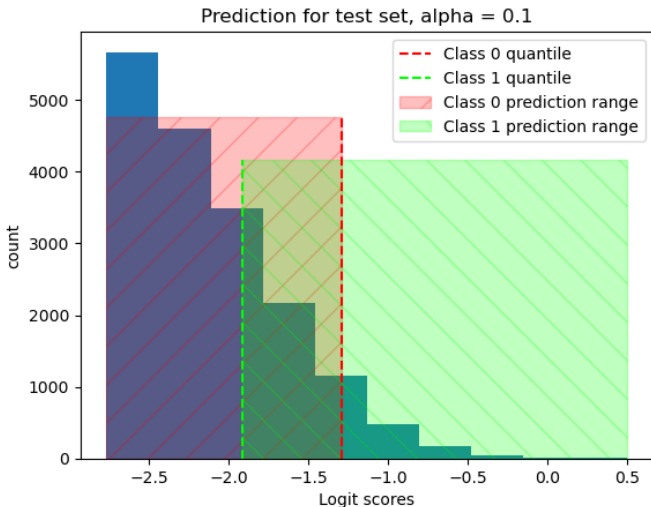
Warunkowa klasyfikacja konformalna - wyniki

Rozkład scorów dla zbioru testowego wyglądał następująco:



Warunkowa klasyfikacja konformalna - obszar klasyfikacji

Przykładowe obszary klasyfikacji. Niektóre obserwacje zostały zaklasyfikowane jednocześnie do klasy 0 i do klasy 1.



Warunkowa klasyfikacja konformalna - pokrycie

Przykładowe pokrycie warunkowej klasyfikacji konformalnej.

1-alpha = 0.9, pokrycie dla klasy 0: 0.9423208580556823

1-alpha = 0.9, pokrycie dla klasy 1: 0.8453237410071942

1-alpha = 0.9, średnie pokrycie: 0.9408064697293047

Druga część zadania 1 polegała na dopasowaniu kwantyli scorów dla segmentacji klientów, czyli chcemy, aby:

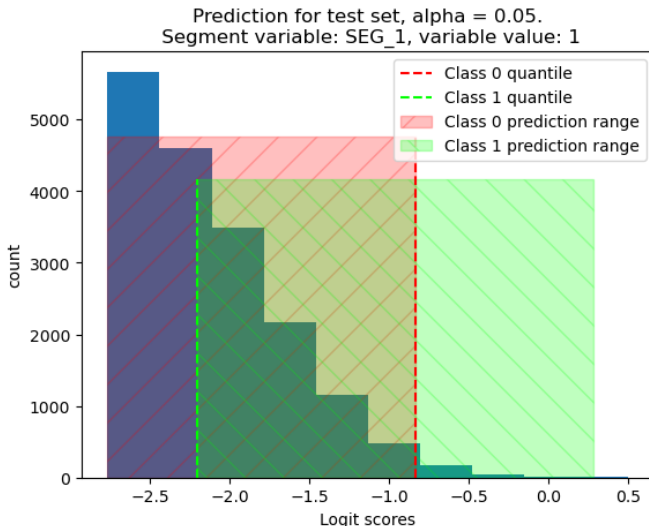
$$\mathbb{P}(Y_{\text{test}} \in \mathcal{C}(X_{\text{test}}) \mid X_{\text{test},1} = g) \geq 1 - \alpha,$$

dla każdej grupy g . W zadaniu grupy wyznaczone zostały z pomocą zmiennych SEG_1 , SEG_2 , SEG_3 .

W tym celu skorzystano z warunkowej klasyfikacji konformalnej.

Segmentacja - obszar klasyfikacji

Przykładowy obszar klasyfikacji dla warunkowej klasyfikacji konformalnej zastosowanej do segmentacji.



Przykładowe pokrycie dla warunkowej klasyfikacji konformalnej zastosowanej do segmentacji, $1 - \alpha = 0.95$.

Zmienna SEG_1, grupa 4, pokrycie dla klasy 0: 0.9692270992366412

Zmienna SEG_1, grupa 4, pokrycie dla klasy 1: 0.9333333333333333

Zmienna SEG_1, grupa 4, średnie pokrycie: 0.9688458815199433

Zadanie 2

Celem zadania 2 było dopasowanie przedziału ufności dla każdego score, wykorzystując prawdopodobieństwo z modelu oraz rzeczywistą klasę.

$$s(x, y) = \frac{|y - \hat{f}(x)|}{u(x)}$$

$$\mathbb{P}[s(X_{\text{test}}, Y_{\text{test}}) \leq \hat{q}] \geq 1 - \alpha \implies \mathbb{P}[|Y_{\text{test}} - \hat{f}(X_{\text{test}})| \leq u(X_{\text{test}})\hat{q}] \geq 1 - \alpha$$

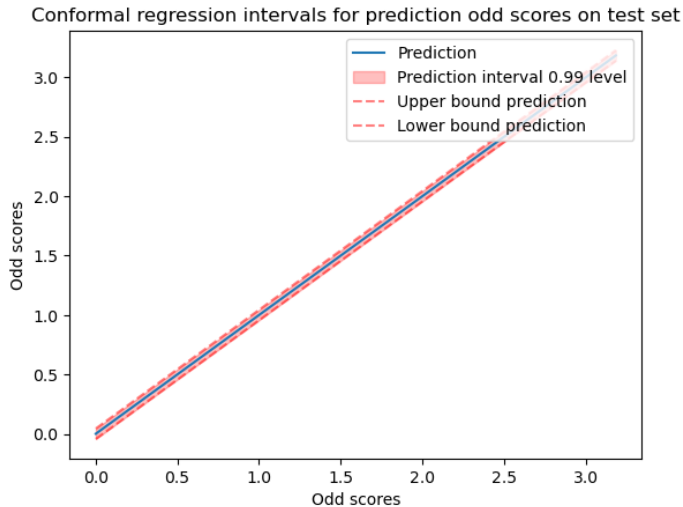
$$\mathcal{C}(x) = [\hat{f}(x) - u(x)\hat{q}, \hat{f}(x) + u(x)\hat{q}]$$

$u(x)$ oznacza pewną miarę niepewności związaną z wartościami zmiennych predykcyjnych, zatem w tym zadaniu przyjęto $\forall_x u(x) = 1$.

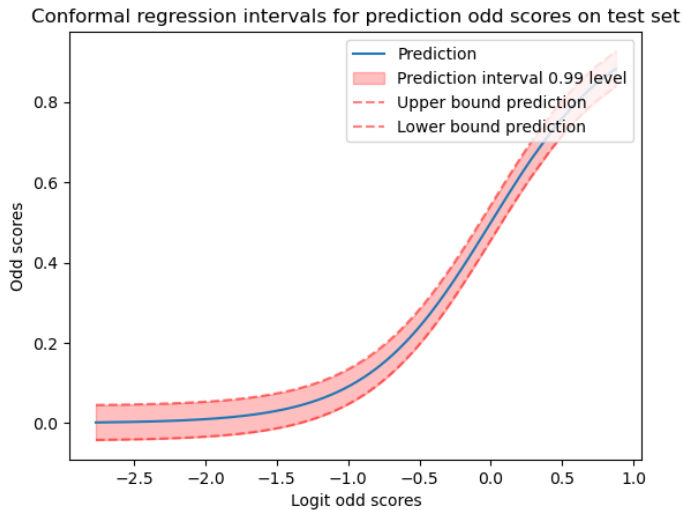
Modelowanym y był iloraz szans $P(Y = 1)/P(Y = 0)$. W celu lepszego określenia kwantyli rozkładu ilorazu szans dokonano estymacji prawdopodobieństwa przynależności do klasy 1:

- dla każdej obserwacji ze zbioru kalibracyjnego wybrano $m = 1000$ obserwacji z najbliższym scorem,
- na podstawie zmiennej TARGET wyznaczono frakcję obserwacji z klasy 1 i wyznaczono estymator ilorazu szans \hat{y} .

Następnie dla każdej obserwacji wyznaczono *conformity scores* $|\hat{y} - \hat{f}(x) / (1 - \hat{f}(x))|$. Ostatecznie wyznaczono kwantyl *conformity scored* rzędu α \hat{q} .



Przedziały ufności - skala logitowa



Zadanie 3

Zadanie 3 jest rozszerzeniem zadania 2. Przedziały ufności powinny zostać skonstruowane dla każdej obserwacji indywidualnie. Aby to osiągnąć, należy wybrać pewną miarę zmienności u związaną z wartościami zmiennych predykcyjnych. Zaimplementowano następujące podejścia:

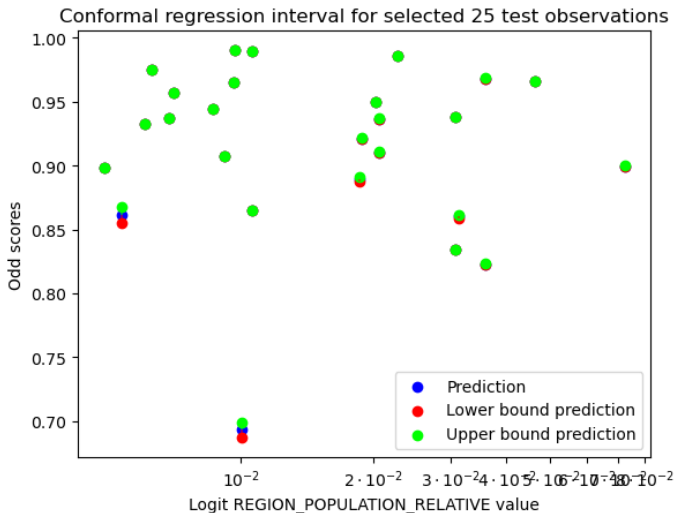
- wariancja predykcji po zaburzeniu wartości początkowej obserwacji,
- wykorzystanie odległości Mahalanobisa.

Przed pokazaniem własnych podejść, podjęliśmy próbę skorzystania z gotowego rozwiązania w pakiecie MAPIE. Oferuje on szereg różnych sposobów wyznaczenia przedziałów predykcji. Niestety, nie udało nam się poprawnie tych metod wykorzystać. Najbardziej podstawowe, które zwracały równe przedziały o tej samej długości dla każdej obserwacji dawały radę się dopasować do danych. Te bardziej skomplikowane zajmowały bardzo długo. Koniecznym było też ograniczenie wielkości zbioru treningowego, ponieważ zbyt duży zbiór powodował błędy kompilacji. Dopiero zbiory po kilka tysięcy wierszy pozwalały na pełną kompilację, ale wtedy traciliśmy dużo danych i wyniki wychodziły znacząco odbiegające od oczekiwanych. Model był dopasowany na $predict_{proba}$ jako skorze.

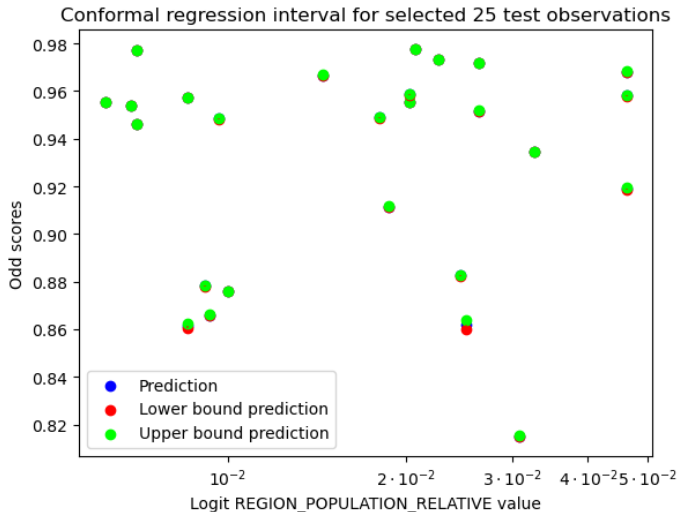
Pierwszym podejściem wykorzystanym w celu wyznaczenia przedziałów konformalnych dla każdej obserwacji było sprawdzenie, w jaki sposób zaburzone dane wpływają na działanie wcześniej przygotowanego modelu. Wykorzystane zostały dane treningowe z pliku `train_kaggle_raw.csv` oraz testowe z pliku `test_kaggle_raw.csv`.

- 1 Model został wytrenowany na początkowych danych, przy użyciu funkcji `predict.proba`, dla każdej obserwacji znaleziono prawdopodobieństwa przynależności
- 2 Niektóre, ciągłe zmienne w danych treningowych zostały zmodyfikowane poprzez dodanie zmiennej normalnej o średniej i wariancji zależnej od danych w tej samej kolumnie. Następnie powtórzony został poprzedni krok (dla zmodyfikowanych danych).
- 3 Drugi krok został powtórzony 10 razy, zwrócona została macierz, która dla każdej obserwacji zawiera 11 prawdopodobieństw przynależności, dla modeli wytrenowanych na różnych zbiorach.
- 4 Funkcję `score` definiujemy jako średnie prawdopodobieństwo przynależności, a wariancja jest wariancją prawdopodobieństw przynależności.
- 5 Analogicznie jak w drugim zadaniu, został wyznaczony kwantyl q , a następnie zgodnie ze wzorem: $[score - var * q, score + var * q]$ zostały wyznaczone przedziały niepewności dla każdej obserwacji.

Przykładowe przedziały dla zbioru testowego



Przykładowe przedziały dla zbioru testowego



Mimo prób zwiększenia wariancji, poprzez stosowanie różnych "szumów", nie udało się nam wystarczająco zwiększyć jej wartości. Skutkuje to bardzo małym przedziałem niepewności, co jest widoczne na poprzednich dwóch grafikach. Wnioskujemy, że model jest odporny na zaburzenia w danych.

Odległość Mahalanobisa

Odległość Mahalanobisa to narzędzie służące głównie do detekcji anomalii (obserwacji odstających/outlierów). Opiera się na prostej idei: im dalej obserwacja znajduje się od środka rozkładu, tym większe wskazanie na bycie outlierem.

Definicja

Niech \mathbb{P} oznacza miarę probabilistyczną na \mathbb{R}^p o średniej $\mu = (\mu_1, \dots, \mu_p)^T$ i macierzy kowariancji Σ . Wtedy odległość Mahalanobisa d obserwacji $X = (X^{(1)}, \dots, X^{(p)})^T$ od μ dla rozkładu \mathbb{P} jest zdefiniowana następująco:

$$d(X, \mu, \Sigma) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)}. \quad (1)$$

Idea wykorzystania odległości Mahalanobisa

Idea stojąca za implementacją odległości Mahalanobisa do powyższego problemu jest następująca. Jeżeli predykcja modelu dla danej obserwacji jest niepewna, to prawdopodobnie obserwacja ta różni się od obserwacji z podobną wartością predykcji.

Schemat postępowania dla odległości Mahalanobisa

Wyznaczanie zmienności dla pojedynczej obserwacji wykorzystuje również podejście konformalne. Najpierw na zbiorze kalibracyjnym wyznaczamy estymowany rozkład odległości Mahalanobisa.

- 1 Dla obserwacji X ze zbioru kalibracyjnego wybierz $m = 1000$ obserwacji z najbliższym scorem.
- 2 Wyznacz wartość oczekiwaną oraz macierz kowariancji obserwacji z wyznaczonej podgrupy.
- 3 Wyznacz odległość Mahalanobisa obserwacji X od średniej wyznaczonej w kroku 2.
- 4 Powtórz kroki 1-3 dla wszystkich bądź wybranych obserwacji ze zbioru kalibracyjnego.

Schemat postępowania dla odległości Mahalanobisa

W kolejnym kroku wyznaczamy odległości Mahalanobisa dla obserwacji ze zbioru testowego oraz porównujemy je z rozkładem "typowych" odległości wyznaczonych na podstawie zbioru kalibracyjnego.

- 1 Dla obserwacji X ze zbioru testowego wybierz $m = 1000$ obserwacji ze zbioru kalibracyjnego z najbliższym scorem.
- 2 Wyznacz wartość oczekiwaną oraz macierz kowariancji obserwacji z wyznaczonej podgrupy.
- 3 Wyznacz odległość Mahalanobisa obserwacji X od średniej wyznaczonej w kroku 2.
- 4 Wyznacz unormowaną rangę ($rank(X)/n_{cal}$) odległości Mahalanobisa dla obserwacji X w wektorze odpowiadającym rozkładowi "typowych" odległości.
- 5 Powtórz kroki 1-4 dla wszystkich obserwacji ze zbioru testowego.

Funkcja u dla metody wykorzystującej odległość Mahalanobisa

Dla obserwacji X wartość zmienności $u(X)$ to $\text{rank}(X)/n_{cal}$ z kroku 3. z poprzedniego slajdu. Oznacza to, że:

- jeżeli odległość Mahalanobisa dla X jest duża względem typowych, czyli $u(X) \approx 1$, to uzyskujemy przedział identyczny jak w zadaniu 2,
- jeżeli odległość Mahalanobisa dla X jest mała względem typowych, czyli $u(X) \approx 0$, to uzyskujemy przedział prawie jednopunktowy.

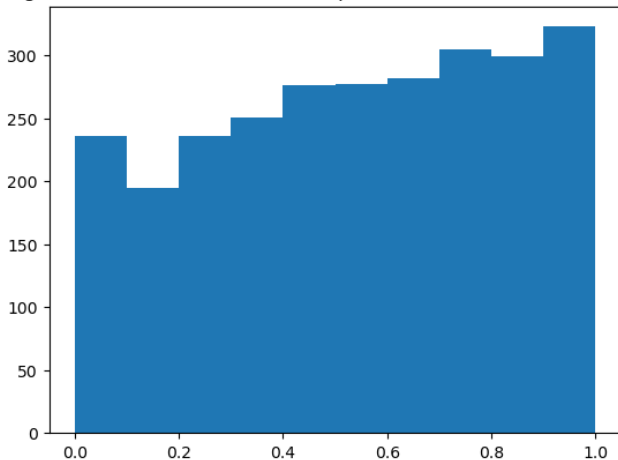
Przyjęcie $u(X) \in (0, 1]$ umotywowane jest naturą problemu: zdecydowana większość wartości predykcji z modelu należą do przedziału $[0, 0.2]$, zatem nawet przyjęcie modelowania ilorazu szans nie sprawi, że przedział modelowanych wartości znacznie się zmieni. Przykładowo, dla zbioru testowego największa wartość ilorazu szans to około 3. Zatem, jeśli przyjęlibyśmy nieograniczoną $u(x)$, to dla niektórych obserwacji przedział ufności zawierałby cały przedział $[0, 1]$.

Uwaga implementacyjna

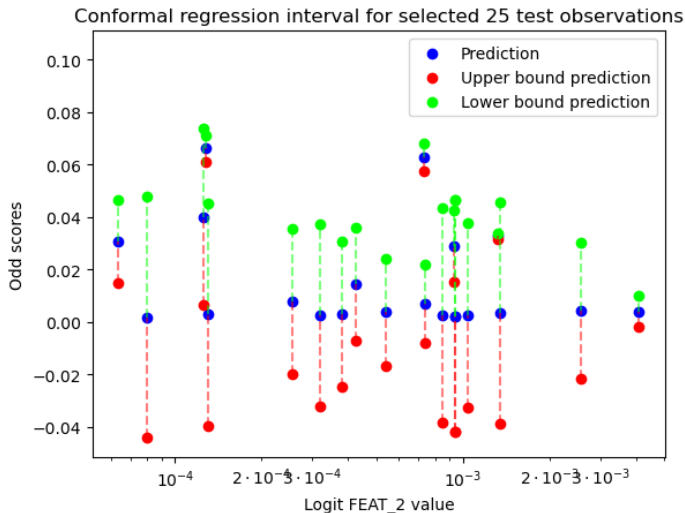
W celu uproszczenia implementacji procedurę z wykorzystaniem odległości Mahalanobisa zastosowano wyłącznie dla obserwacji, które nie miały braków danych. Możliwe jest rozszerzenie powyższego schematu o uzupełnianie braków danych i uwzględnienie liczby braków danych w odległości Mahalanobisa bądź w funkcji u .

Rozkład rang dla zbioru testowego

Histogram of variance of test set samples, based on Mahalanobis Distance



Przykładowe przedziały dla zbioru testowego



Przykładowe przedziały dla zbioru testowego

