

Sprawozdanie nr. 2

1. Tytuł

Regresja liniowa, regresja logistyczna

Hubert Król

2. Wstęp

a) Cel ćwiczenia

Przeprowadzenie regresji liniowej na dataset nr 5.
Synchronous Machine.

Data set link = <https://archive.ics.uci.edu/ml/datasets/Synchronous+Machine+Data+Set>

Zmieniając współczynnik nauczania lub początkowe wagi i wyraz wolny dla regresji, będę obserwował zmiany zachodzące w wyliczonej dokładności porównując do domyślnego modelu.

b) Niezbędne podstawy teoretyczne

W regresji liniowej opieramy naszą prognozę na poniższym działaniu:

Wzór 1. Prognozowana wartość dla n liczby argumentów

$$y_prediction = \theta_n x_n + \theta_{n-1} x_{n-1} + \theta_{n-2} x_{n-2} + \dots + \theta_2 x_2 + \theta_1 x_1 + \theta_0$$

$y_prediction$ – wartości prognozowane

θ_n – współczynnik dla argumentu

x_n - argument

Działanie na obliczanie Miary błędu:

Wzór 2 koszt

$$cost = \frac{1}{2m} \sum [(y_{prediction} - Y)^2]$$

cost - koszt regresji

m – liczba argumentów

$y_prediction$ – wartości prognozowane

Y – wartość docelowa

Numeryczny sposób na obliczenie zmiany wag i wyrazu wolnego:

Wzór 3 spadek gradientu

$$d_theta = \frac{1}{m}(X^T * (y_prediction - Y))$$

Wzór 4

$$theta = theta - \alpha * d_theta$$

α - współczynnik nauczania

d_theta – spadek gradientu

m – liczba wierszy

X^T – macierz transponowana danych wejściowych

y_prediction – wartości prognozowane

Y – wartość docelowa

- a) Wybieramy parametry X, Y, szybkość uczenia się (α) i iteracje.
- b) Definiujemy m jako rozmiar zbioru
- c) Theta będzie naszym wektorem zer.
- d) Uruchamiamy pętlę i w każdej iteracji przeprowadzamy 4 powyższe równania
- e) Śledzimy nasz koszt przy każdej iteracji
- f) Na koniec zwracamy parametr theta oraz listę kosztów

3. Wymagania techniczne

Algorytm został opracowany na:

- Anaconda Navigator (anaconda 3) Spyder (Python 3.9)
- System operacyjny: Windows 10
- plik zaczytujący dane z pliku dane.xlsx
- Użyte biblioteki: numpy, pandas, matplotlib, sklearn
- plik oryginalny .csv przekonwertowałem na.xlsx ponieważ dane były zapisane w kolumnach

Nie po przecinkach

4. Przedstawienie Wyników

Wykres kosztu do iteracji

a) Iteracje 100, współczynnik uczenia 0.0005

Wykres 1:

Koszt



Iteracje

Obraz 1 a)

```
Error is : 23.790148670216084 %  
Accuracy is : 76.20985132978392 %
```

Obraz 1 b)

```
Cost is : 1.1858631887640447
Cost is : 0.6127652363668856
Cost is : 0.32439531183310205
Cost is : 0.17928466493644946
Cost is : 0.10625411534380028
Cost is : 0.06949027170915775
Cost is : 0.0509738340799661
Cost is : 0.04163849627497135
Cost is : 0.03692259894748011
Cost is : 0.03453096965145613
```

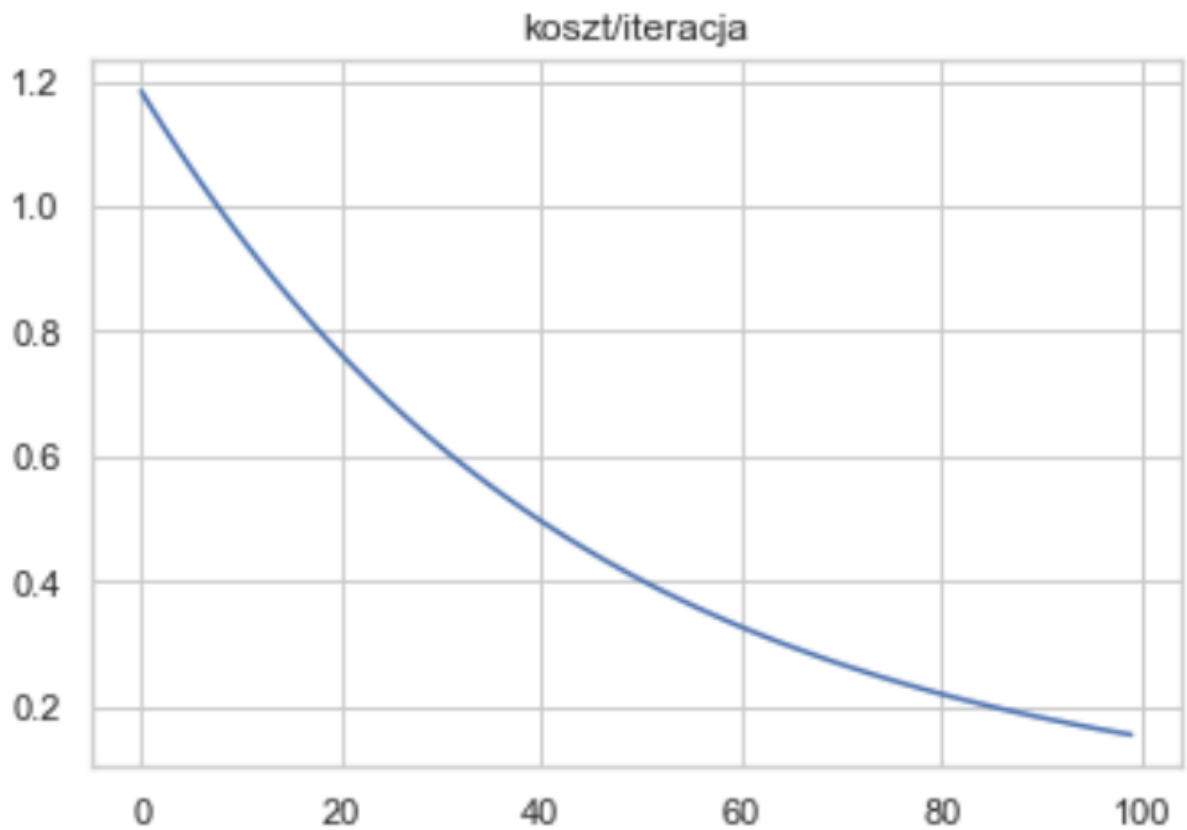
Error – różnica między wynikiem docelowym oraz prognozowanym 23%

Accuracy – zgodność danych przewidywanych na 76%

b) Iteracje 100, współczynnik uczenia 0.0005.

Wykres 2

Koszt



Iteracje

Obraz 2 a)

```
Error is : 23.001690110560315 %  
Accuracy is : 76.99830988943968 %
```

Obraz 2 b)

```
Cost is : 1.1807745775280898  
Cost is : 0.6128136432978961  
Cost is : 0.3258228562073975  
Cost is : 0.18079663324656742  
Cost is : 0.10750024086894783  
Cost is : 0.07044648717831817  
Cost is : 0.05170491631058733  
Cost is : 0.04221588643523125  
Cost is : 0.03740187252718772  
Cost is : 0.034950008811611755
```

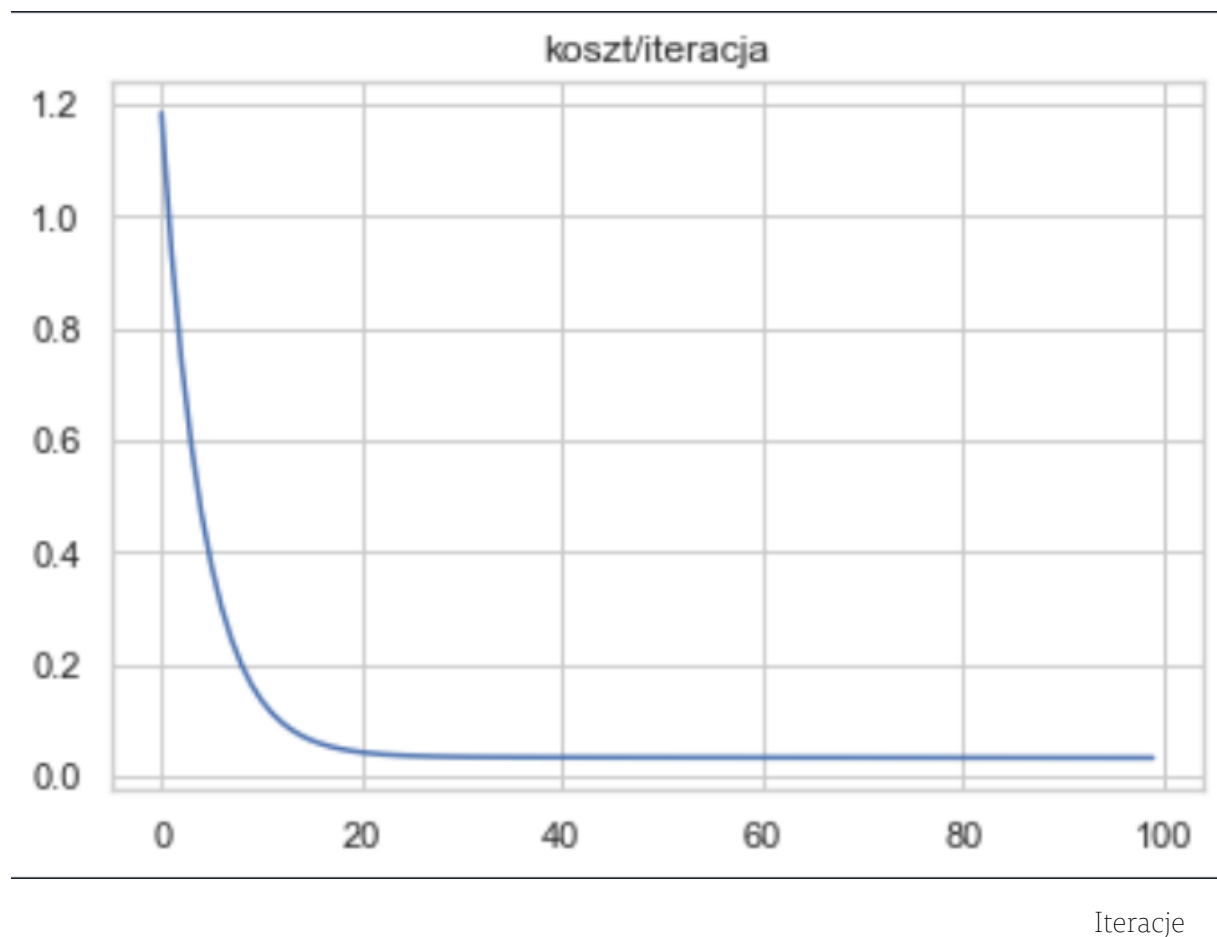
Error – różnica między wynikiem docelowym oraz prognozowanym 23%

Accuracy – zgodność danych przewidywanych na 76%

c) Iteracje 100, współczynnik uczenia 0.005

Wykres 3:

Koszt



Obraz 3 a)

```
Error is : 19.97133604399402 %  
Accuracy is : 80.02866395600599 %
```

Obraz 4 a)

```
Cost is : 1.1826439359550565  
Cost is : 0.0343219094455736  
Cost is : 0.03312552261651549  
Cost is : 0.03271648998853179  
Cost is : 0.03231334533627399  
Cost is : 0.0319154802319856  
Cost is : 0.03152281968096466  
Cost is : 0.031135290202138607  
Cost is : 0.030752819446208734  
Cost is : 0.030375336176457598
```

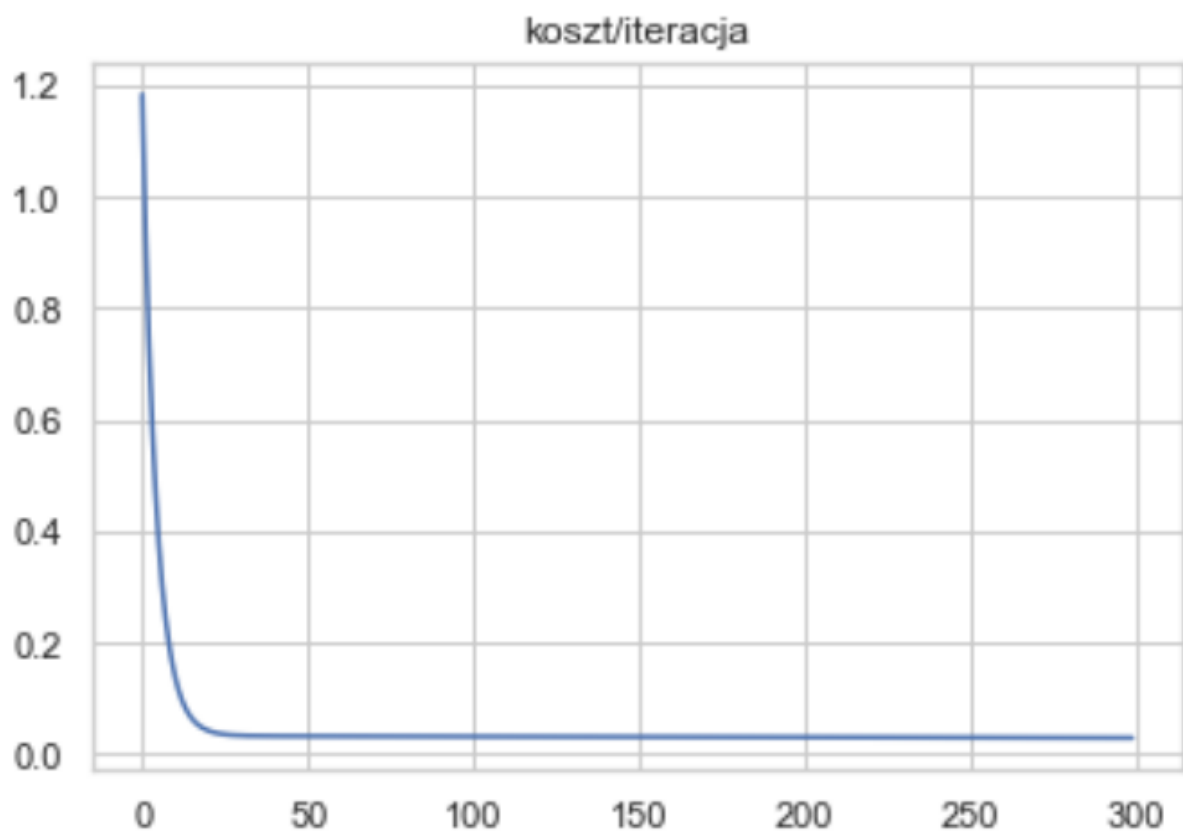
Error – różnica między wynikiem docelowym oraz prognozowanym 19%

Accuracy – zgodność danych przewidywanych na 80%

d) Iteracje 300, współczynnik uczenia 0.005.

Wykres 4:

Koszt



Iteracje

Obraz 4 a)

```
Error is : 20.460478001072385 %  
Accuracy is : 79.53952199892763 %
```

Obraz 4 b)

```
Cost is : 1.183417982022472  
Cost is : 0.03344683751632553  
Cost is : 0.03220016319665199  
Cost is : 0.03180455573128145  
Cost is : 0.03141484502605646  
Cost is : 0.031030330843002833  
Cost is : 0.030650935945968287  
Cost is : 0.030276584818074  
Cost is : 0.029907203181655684  
Cost is : 0.029542717975687645
```

Error – różnica między wynikiem docelowym oraz prognozowanym 20%

Accuracy – zgodność danych przewidywanych na 79%

5. Podsumowanie

- a) Im większy współczynnik uczenia tym większa zgodność
- b) Im mniejszy współczynnik uczenia tym większy tym mniejsza szansa na wystąpienie błędu
- c) Im mniejszy współczynnik uczenia i większa ilość iteracji uczenia tym dokładniejsze wartości
- d) Im większa ilość iteracji i większy współczynnik uczenia tym wykres kosztu do iteracji staje się bardziej „stromy” w przeciwnym wypadku kiedy mamy mniejszy współczynnik uczenia wykres jest bardziej się „prostuje”

6. Bibliografia

- https://drive.google.com/file/d/1F9U08jxVZScPgRT3kHl8GhN1lMvf_gl-/view
- zmo_data_nd.pdf

Opracowanie: Hubert Król