

GOLEM

KOŁO NAUKOWE SZTUCZNEJ INTELIGENCJI

Regresja liniowa i zagadnienie redukcji wymiarów

Antoni Kowalczyk

KNSI GOLEM BOOTCAMP #2

A large, dark blue ink splash or blotch serves as the background for the text. It has irregular, organic edges with some lighter blue and white speckling around it, giving it a textured, artistic appearance.

Regresja liniowa

Zarys problemu



Czym jest regresja?

„Mając dane wejściowe zwróć ciągłą wartość”

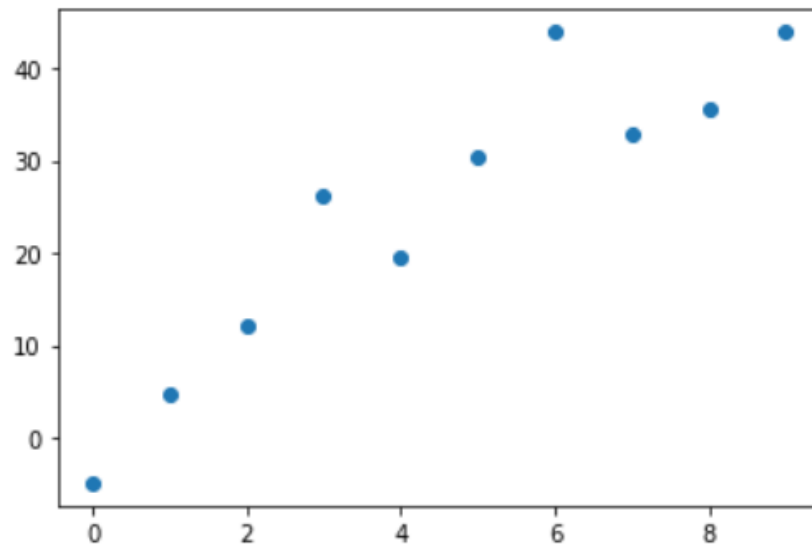
Kolejne z głównych zagadnień Machine Learning’owych (po klasyfikacji)

Polega na przybliżeniu wyjścia funkcją liniową wielu zmiennych i wielu parametrów

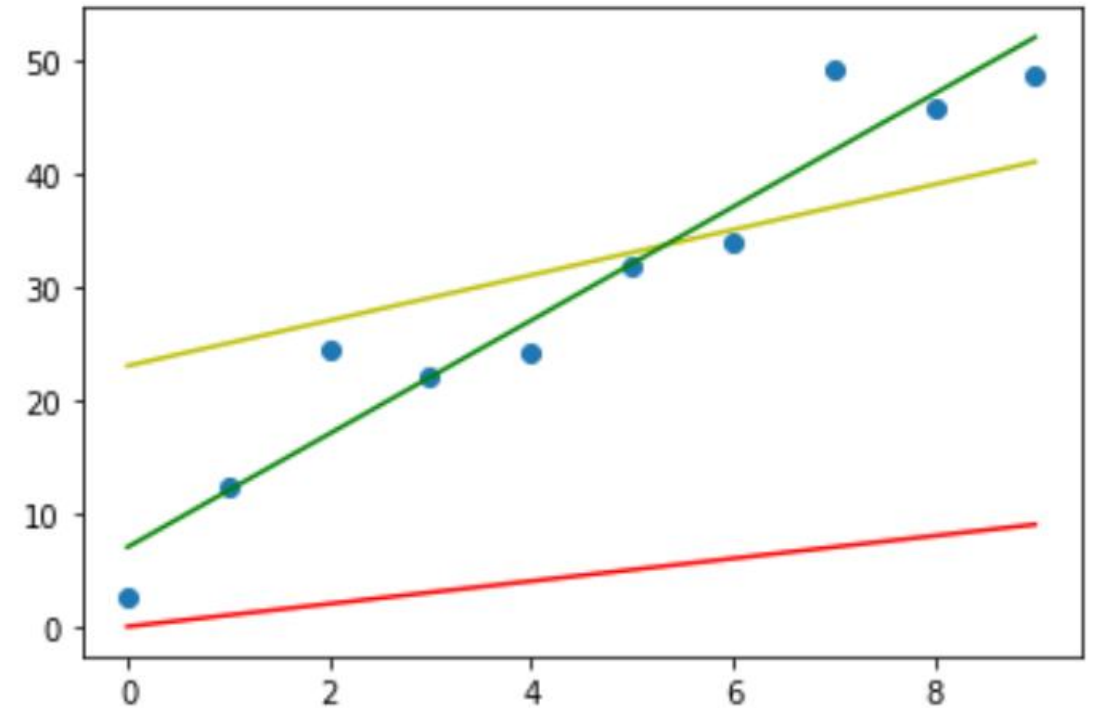
Przykład


Dane: **Wizualizacja tych danych:**

x	y
0	-4.9
1	4.6
2	12.2
3	26.3
4	19.4
5	30.5
6	43.9
7	32.9
8	35.6
9	43.9



Przykładowe przybliżenia y funkcją liniową:





Jak wybrać
najlepszą
aproksymację?

Funkcja „straty”

Zagadnienie minimalizacji funkcji straty


Jak zminimalizować funkcję?

Funkcja straty

Metryka pozwalająca nam porównywać modele

Dla regresji liniowej najpopularniejsza to MSE (Mean squared error)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$



Minimalizacja funkcji straty

Po co: po to, aby nasz model przybliżał rzeczywistość w najlepszy możliwy sposób

Jak możemy to osiągnąć: metodami optymalizacyjnymi


Metody minimalizacji

Metoda najmniejszych kwadratów

- Daje precyzyjne i szybkie dla małych zbiorów rozwiązanie
- Dla dużych ilości danych/innych funkcji strat jednak działa słabo
- Praktycznie nie używane

Metoda gradientowego spadku

- Jeśli estymujemy gradient, to wynik może znaleźć się dowolnie blisko minimum
- Szybka metoda dla dowolnych funkcji strat (różniczkowalnych)
- Używana powszechnie (może nie w tej konkretnej formie)



Metoda gradientowego spadku

Jak działa?

Dlaczego działa?

Wyjaśnienie

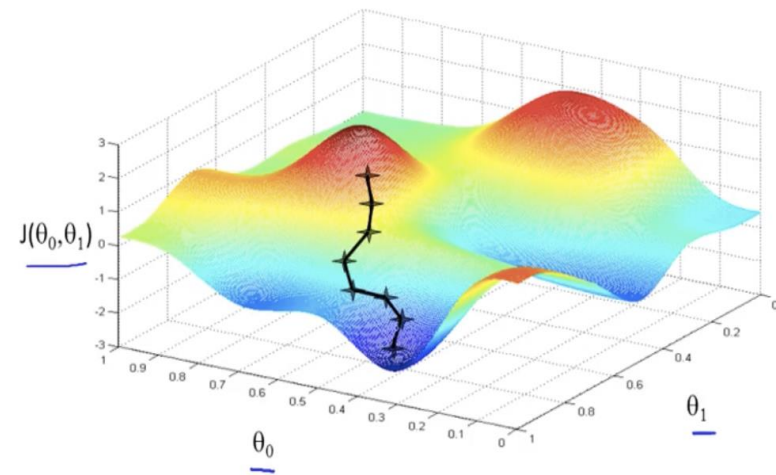
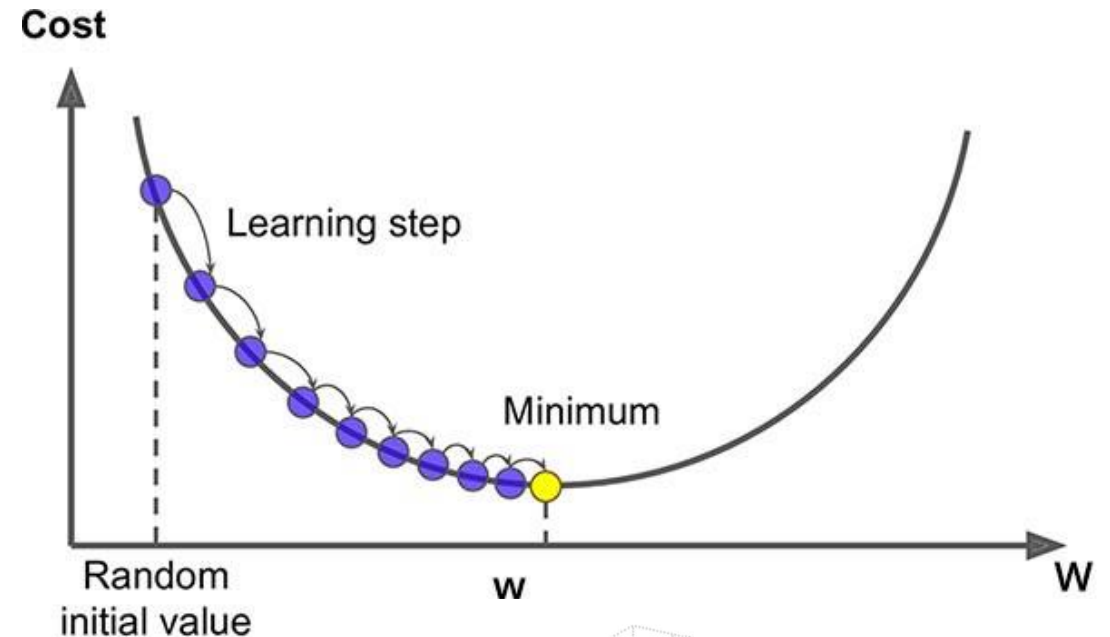
Czym jest gradient?

Jest to wektor pochodnych

Dlaczego jest użyteczny?

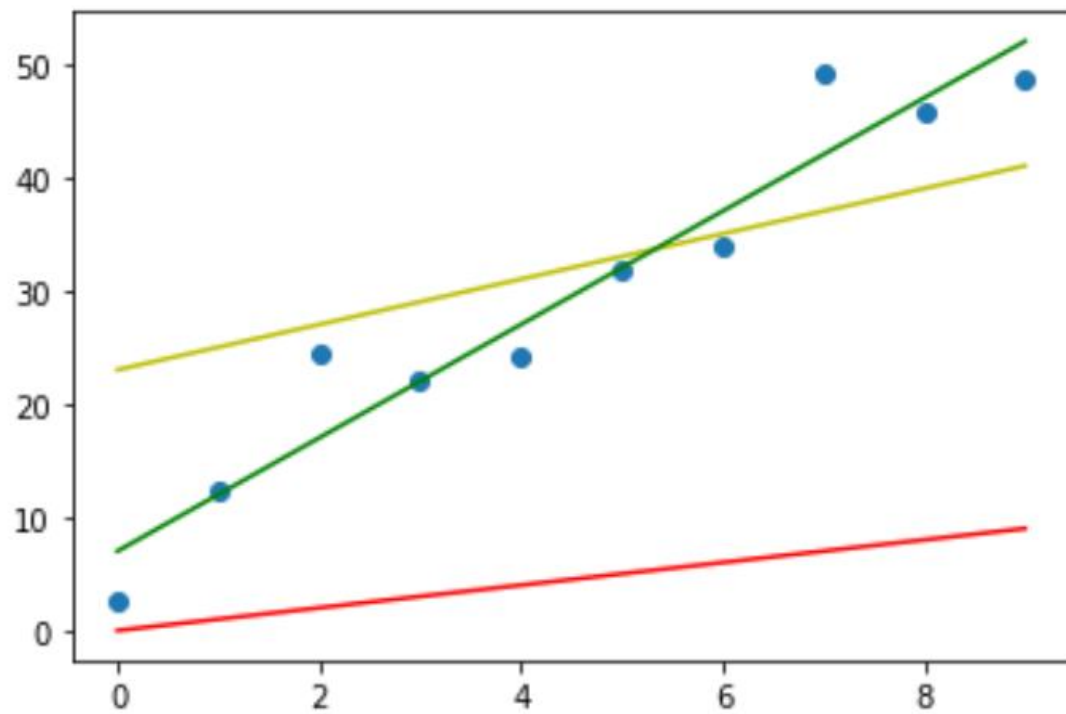
Wskazuje nam kierunek najszybszego wzrostu funkcji → poruszając się przeciwnie do tego kierunku zmierzamy ku minimum funkcji

Wizualizacja



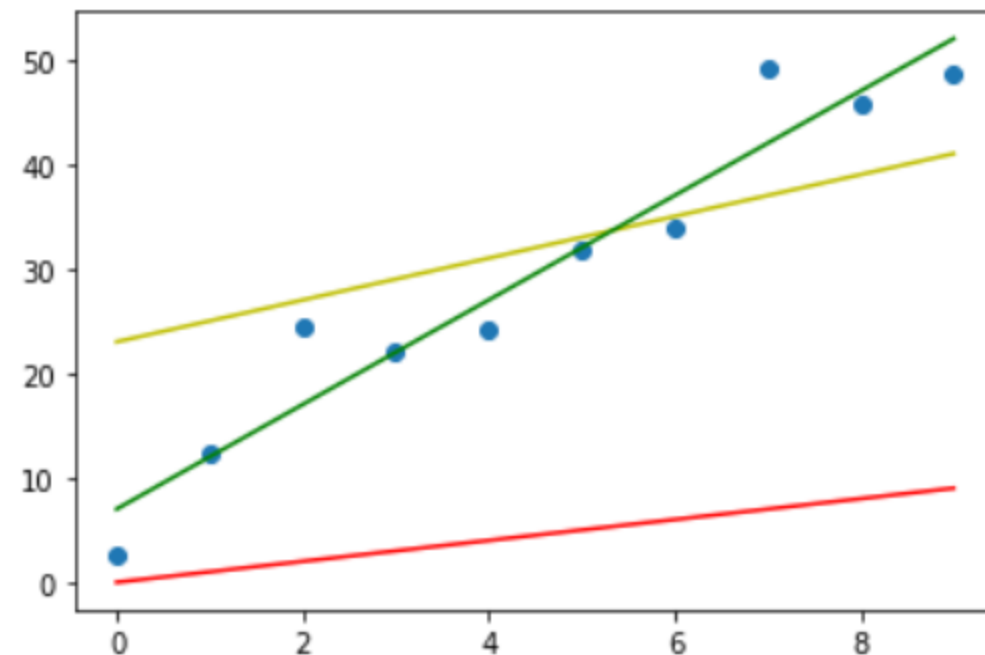
Przykład (c. d.)

Przykładowe przybliżenia y funkcją liniową:



MSE tychże przybliżeń:

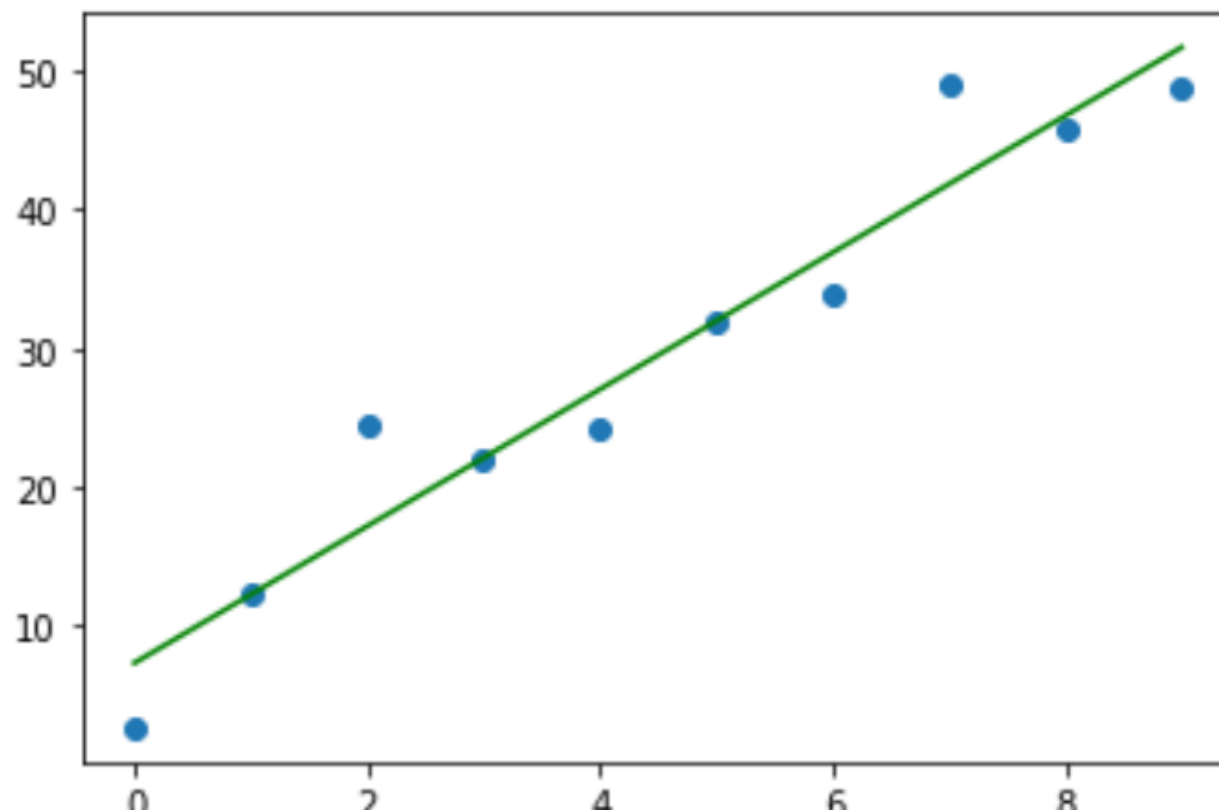
red line MSE: 769.22
yellow line MSE: 92.26
green line MSE: 15.48



Najlepsza funkcja liniowa na tych danych

`Sklearn LinearRegression()` MSE: 15.44


Our best "model" MSE: 15.48





Przejdźmy do
praktyki :)

Przykład na prawdziwym datasetcie



Redukcja wymiarowości

W skrócie



Czym to jest i
po co to nam?

Jest to transformacja wysoko-
wymiarowej przestrzeni naszych danych
w nisko-wymiarową przestrzeń

Zależy nam na tym, żeby dokonując tej
transformacji utracić minimalną ilość
danych

Chcemy zredukować wymiar problemu,
żeby szybciej uczyć nasze modele

PCA (Principal Component Analysis)

Jest to jedna z najbardziej popularnych metod redukcji wymiarowości

Jak to działa? (w dużym skrócie, matma dla ciekawych [TUTAJ](#))

Przykład w kodzie