

DATA MINING 5

Zadanie 1

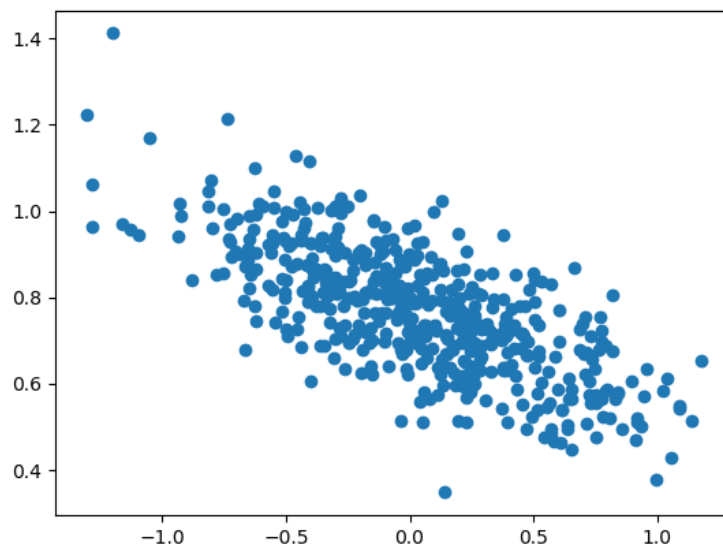
1. Policz gradient (pochodną) funkcji $f(x) = 2x^2 + 3x + 4$ w punkcie 4.
2. Policz gradient funkcji $f(x, y) = 2x^3 + 3y^2 + 4$ w punkcie (4,5).

Zadanie 2

Sprawdź dla funkcji z punktu 2 (wykonując odpowiednie rachunki), że jeżeli **przesuniemy się w kierunku przeciwnym do gradientu** o pewną niewielką wartość to wartość funkcji $f(x, y) = 2x^3 + 3y^2 + 4$ zmaleje.

Zadanie 3

Wygeneruj **zbiór treningowy** o następującym kształcie:



- A. Zastosuj **regresję liniową** i znajdź prostą aproksymującą.
- B. Przetestuj działanie modelu dla **różnej ilości epok** (np. 30,60,100,200,500) i **różnych wartości współczynnika uczenia** (np. 0.001, 0.01, 0.1, 0.6, 1.0). W każdym przypadku wygeneruj wykres zmian błędu.
- C. Dla jakich wartości **hiperparametrów** uzyskana ostatecznie wartość błędu jest najmniejsza, a dla jakich największa.

Zadanie 4

Zastosuj **regresję liniową** i znajdź prostą aproksymującą dla atrybutów **rm** i **medv** ze zbioru danych **Boston.csv**. Przetestuj działanie modelu dla **różnej ilości epok** i **różnych wartości współczynnika uczenia**.

Zadanie 5

Zastosuj **regresję logistyczną** do problemu z dwoma gangami **obliczając wprost gradient funkcji błędu**. Wykorzystaj następującą definicję (tzn. entropia krzyżowa):

```
def loss_fn(y, y_model):  
    return tf.reduce_mean(-y*tf.math.log(y_model) -  
                           (1-y)*tf.math.log(1-y_model))
```

oraz funkcję `tf.sigmoid`.

- A. Sprawdź działanie modelu dla punktów: (3,2), (7,6), (5,4)
- B. Przetestuj działanie modelu dla kilku różnych wartości hiperparametrów: **ilość epok** i **współczynnik uczenia**.
- C. Dla jakich wartości hiperparametrów uzyskałeś najlepszy/najgorszy wynik?

Zadanie 6

Zastosuj **regresję softmax** do problemu z dwoma gangami. **Wykorzystaj bibliotekę keras**.

- A. Przyjmij podział zbioru danych: **80%** - dane **treningowe**, **20%** - dane **walidacyjne**.
- B. Przetestuj działanie modelu dla **różnych wartości hiperparametrów**. Dla jakich wartości uzyskałeś najlepszy/najgorszy wynik?
- C. Wygeneruj wykres zmiany błędu dla **danych treningowych i walidacyjnych**.

Wynik: notatnik Colab z rozwiązaniem + jego wydruk PDF oraz omówienie wyników.

Zadanie 7*

Zastosuj **regresję softmax** do problemu z dwoma gangami. **Gradient funkcji błędu oblicz wprost**.

- A. Parametry zapisz w macierzy i do wyliczenia wartości zwracanych przez model wykorzystaj operację **tf.matmul**. Wykorzystaj:

```
w = tf.Variable(np.random.random((2, 2)))  
b = tf.Variable(np.random.random((2)))
```

- B. Wykorzystaj funkcję błędu:

```
loss_fn = tf.keras.losses.BinaryCrossentropy()
```

- C. Przetestuj działanie programu dla kilku różnych wartości parametrów: **ilość epok**, **współczynnik uczenia**, **wielkość batch**. Dla jakich wartości wynik jest najlepszy/najgorszy.
- D. Przyjmij podział zbioru danych: **80%** - dane **treningowe**, **20%** - dane **walidacyjne**.