ОСНОВИ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ, НЕЙРОННИХ МЕРЕЖ та ГЛИБОКОГО НАВЧАННЯ

Модуль 2. Навчання з вчителем

Лекція 2.4.
Приклад вирішення задачі поліноміальної регресії за допомогою Scikit-learn

Регресія

Формально. Маємо:

Незалежні змінні ⇔ $\boldsymbol{X}$

Залежна змінна ⇔ $\boldsymbol{Y}$

Відомі пари ⇔ $(\boldsymbol{x}_i, \hat{y}_i),\ i = 1,2,\dots,m$

Регресійна функція (модель) ⇔ $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{W}, \boldsymbol{X})$, де $\boldsymbol{W}$ − множина невідомих параметрів $w_0, w_1, \dots w_K$.

Необхідно ⇔ оцінити унікальні значення $\boldsymbol{W}$, які в деякому розумінні підходять **найкраще**, а регресійна модель у застосування до даних може розглядатися як перевизначена система для $\boldsymbol{W}$.

Інакше, знайти такі значення $w_0, w_1, \dots w_K$, які мінімізують

$$\boldsymbol{L} = \sum_1^m \boldsymbol{\varepsilon}_i^2 \text{ , де } \boldsymbol{\varepsilon}_i = \widehat{\boldsymbol{y}}_i - \boldsymbol{y}_i = \widehat{\boldsymbol{y}}_i - \boldsymbol{f}(\boldsymbol{W}, \boldsymbol{x}_i)$$

Лінійна регресія

Регрессія

- **Лінійна регресія**
- **Поліноміальна регресія**
- **. . . .**
- **Нелінійна регресія**

$$\boldsymbol{f(W, X) = w_0 + w_1 X_1^D + w_2 X_2^D \ldots}$$

Вектор *D* - зведення в ступінь (збігається з розмірністю *X*)

Оцінка якості за допомогою квадратичної помилки

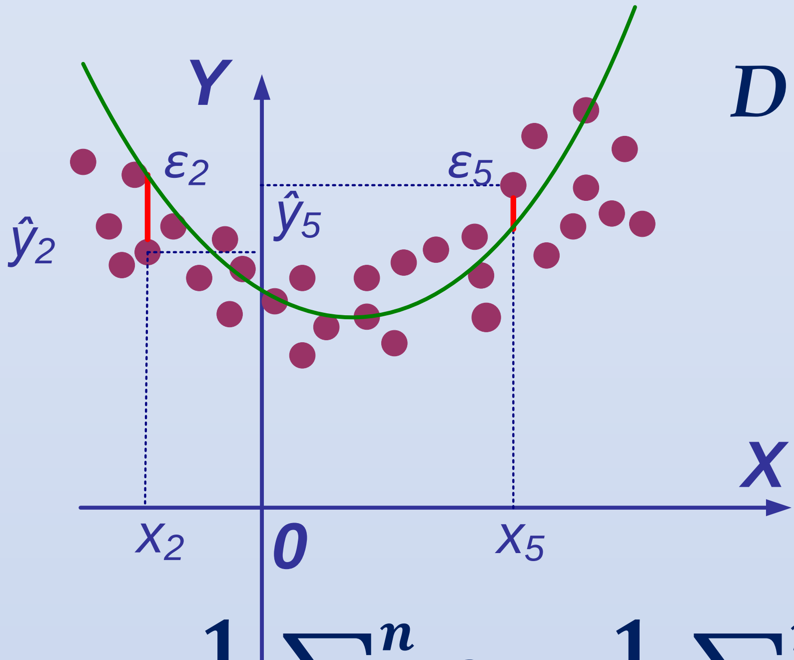
$$\boldsymbol{MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}$$

Найпростіша поліноміальна регресія

x_1, y_1
x_2, y_2
$\ldots\ldots$
x_n, y_n

$$\hat{\mathbf{y}} = f(w_0, w_1, x) = w_0 + w_1 x + w_2 x^2 + \cdots$$

D – скаляр = 2



$$L(w) = \frac{1}{2n}\sum_1^n \varepsilon_i^2 = \frac{1}{2n}\sum_1^n (\hat{y}_i - y_i)^2 =$$

$$\frac{1}{2n}\sum_1^n \left(\hat{y}_i - (w_0 + w_1 x + w_2 x^2)\right)^2$$

Знайти такі $(\overline{w})$, що $L(\overline{w}) = \min_w L(w)$

Множинна поліноміальна регресія

$X = \begin{bmatrix} x_1 \\ \dots \\ x_m \end{bmatrix},$ D – скаляр = 2

$\hat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{W}, \boldsymbol{X})$

X_1, y_1

X_2, y_2

$\dots\dots$

X_n, y_n

$$= \boldsymbol{W_0} + \langle \boldsymbol{W_1 X} \rangle + \langle \boldsymbol{W_2 X^2} \rangle + \sum \boldsymbol{w_{3_{i,j}} * x_i * x_j}$$

$\boldsymbol{W_q}$ – вектори ваг відповідної форми

Зауважте : $\hat{\boldsymbol{y}}$ лінійно залежить від ваг!

$$\boldsymbol{L}(\boldsymbol{w}) = \sum_{1}^{n} \boldsymbol{\varepsilon_i^2} = (\hat{y}_1 - \boldsymbol{y_1})^2 + \dots + (\hat{y}_i - \boldsymbol{y_i})^2 + \dots$$

Знайти такі $(\overline{\boldsymbol{W}})$, **що** $\boldsymbol{L}(\overline{\boldsymbol{W}}) = \min_{\boldsymbol{w}} \boldsymbol{L}(\boldsymbol{W})$

Лінійна регресія з scikit-learn

Бібліотека *scikit-learn*

Модуль *preprocessing*

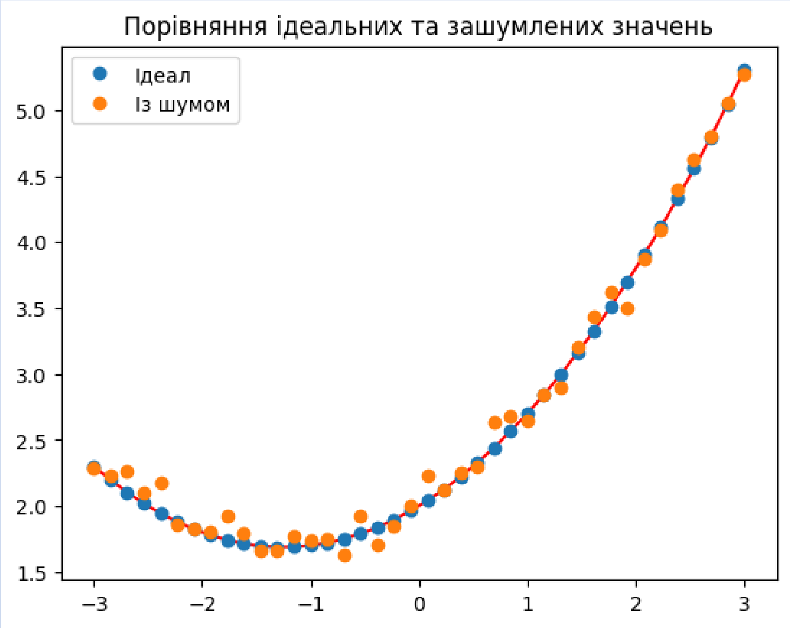
Клас *PolynomialFeatures*

PolynomialFeatures(degree=2, *, interaction_only=False, include_bias=True, order='C')

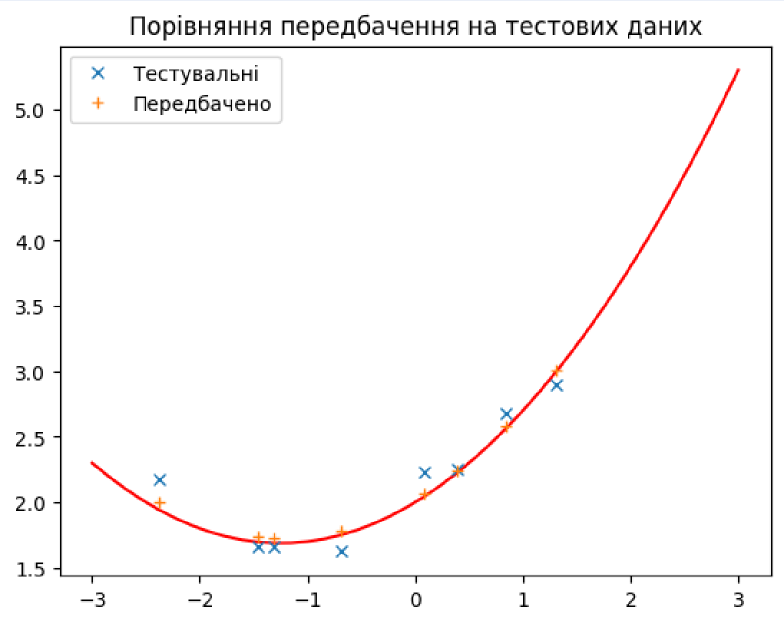
Створює нову матрицю ознак, що складається з усіх поліноміальних комбінацій ознак зі ступенем, меншим або рівним зазначеному **degree**.

Результати

Приклад - lec_02_04_Exmpl_1

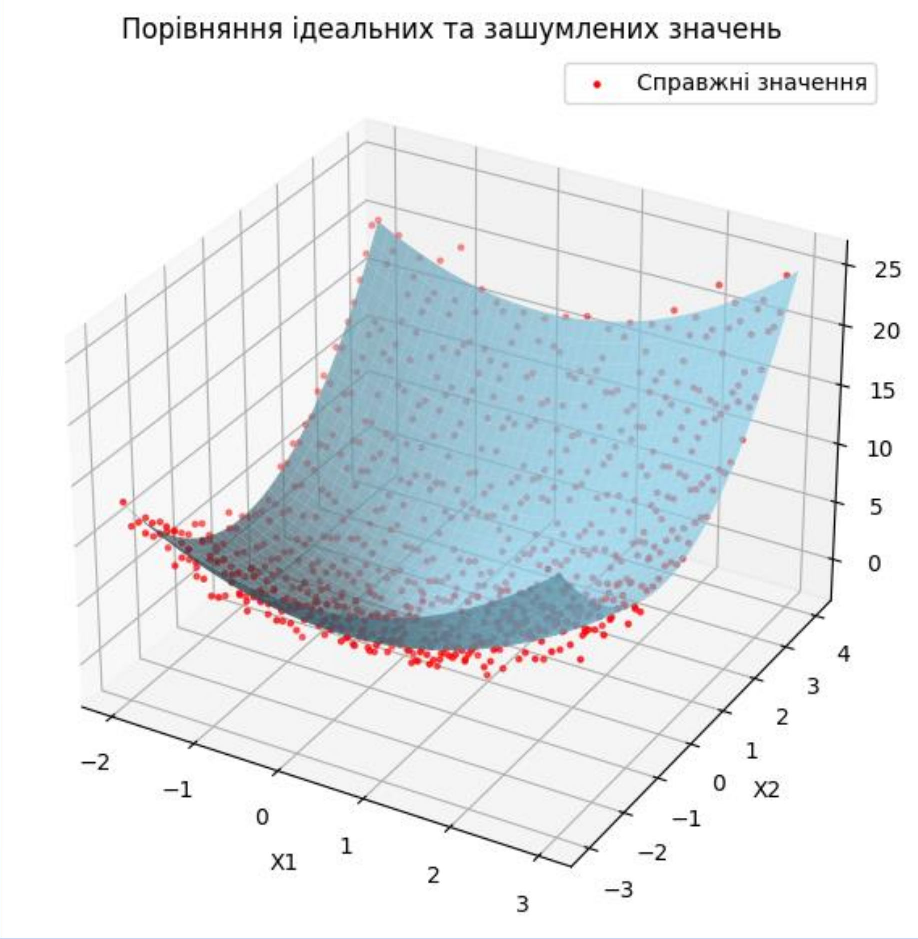


MSE = 0.01

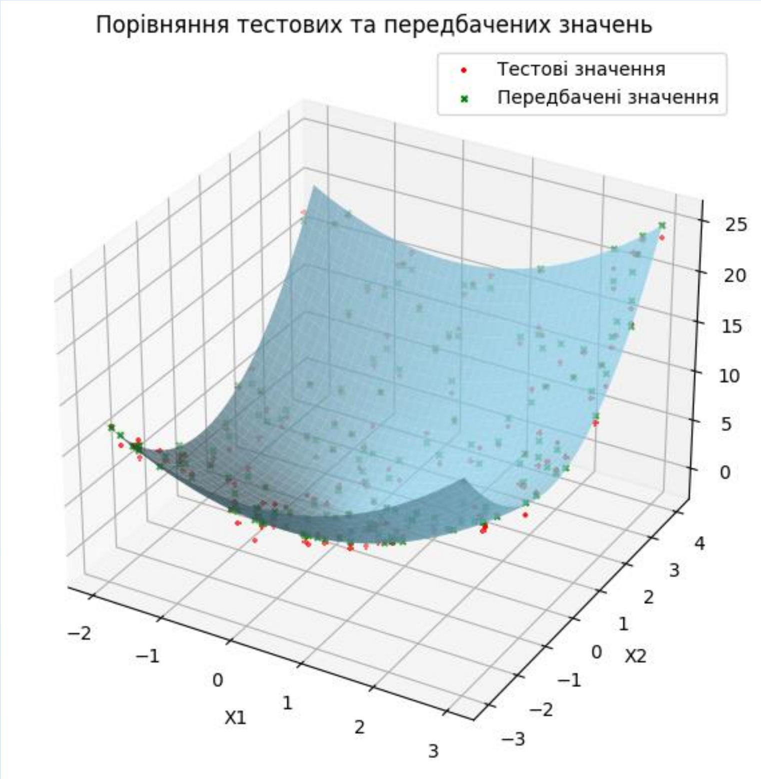


Результати

Приклад - lec_02_04_Exmpl_1



MSE = 1.0



Контрольні запитання

- **Надайте загальну постановку задачі регресії.**
- **Вкажіть особливості задачі поліноміальної регресії.**
- **Поясніть сутність методу вирішення задачі поліноміальної регресії за допомогою трансформації до лінійної.**
- **Наведіть приклад застосування *PolynomialFeatures* бібліотеки *scikit-learn* для вирішення задачі поліноміальної регресії.**

Рекомендована ЛІТЕРАТУРА

- **Глибинне навчання:** Навчальний посібник / Уклад.: В.В. Литвин, Р.М. Пелещак, В.А. Висоцька В.А. – Львів: Видавництво Львівської політехніки, 2021. – 264 с.
- Тимощук П. В., Лобур М. В. **Principles of Artificial Neural Networks and Their Applications: Принципи штучних нейронних мереж та їх застосування:** Навчальний посібник. – Львів : Видавництво Львівської політехніки, 2020. – 292 с.
- Morales M. **Grokking Deep Reinforcement Learning. –** Manning, 2020. – 907 с.
- Trask Andrew W. **Grokking Deep Learning.** – Manning, 2019. – 336 с.

The END
Модуль 2. Лекція 04.