Regresja liniowa - to jedna z podstawowych metod statystycznych i uczenia maszynowego, która służy do modelowania zależności między zmiennymi. Regresji używa się, aby przewidzieć wartość zmiennej zależnej (cecha ilościowa)

Wzór ogólny:

wektor wag

$$\hat{y}(\beta, x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N$$

przewidywana zmienna/cecha

zmienne niezależne (posiadane dane)

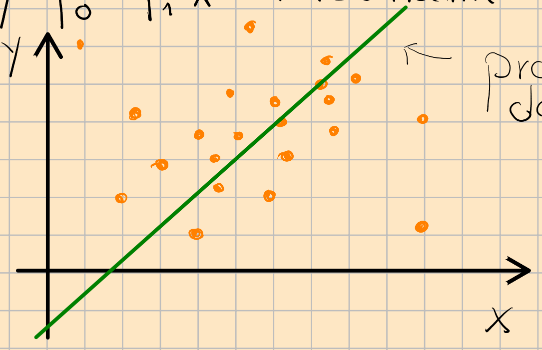
wynik $\hat{y}$ zależy od danych wejściowych x oraz od wag β

$$\beta = [\beta_0, \beta_1, \beta_2, \dots, \beta_N]$$
$$X = [X_1, X_2, X_3, \dots, X_N]$$

! Uwaga:
posiadane dane wejściowe nie ulegają zmianie - są to wartości stałe. W celu poprawy modelu można zmieniać tylko wektor wag.

Przypadek 2D:

$\hat{Y} = \beta_0 + \beta_1 X$ ← równanie prostej

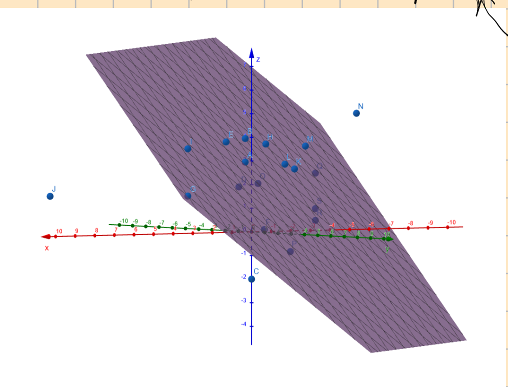


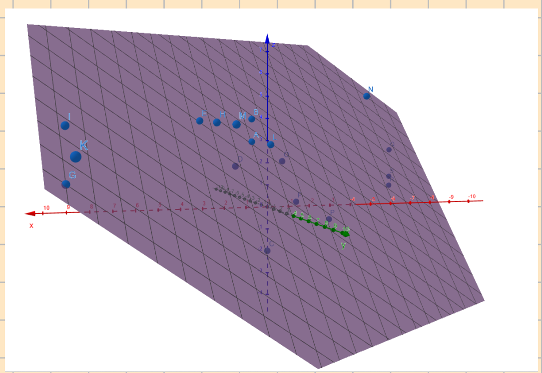
prosta, która najlepiej przybliża chmurę danych, inaczej przestrzeń danych

Przypadek 3D:

$\hat{y} = \beta_0 + \beta_1 x + \beta_2 y$ → inaczej → $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x$

← równanie płaszczyzny





Przypadek 4D:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$ ← równanie hiperpłaszczyzny w przestrzeni 4D

zmienna zależna y jest opisana poprzez trzy zmienne niezależne

Ponieważ regresja liniowa opisuje zależność liniową, to hiperpłaszczyzna jest „płaska" w danym wymiarze.

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources	Extracurricular_Activities	Sleep_Hours
0	23	84	Low	High	No	7
1	19	64	Low	Medium	No	8
2	24	98	Medium	Medium	Yes	7
3	29	89	Low	Medium	Yes	8
4	19	92	Medium	Medium	Yes	6

. . . . .

Gender	Exam_Score
Male	67
Female	61
Male	74
Male	71
Female	70

x^j ← konkretny wiersz ma dalej oznaczenie x^j

x_2^j

Wartość na drugiej współrzędnej „wektora" x^j

Zmienne niezależne

Zmienna zależna

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{Hours\_Studied} + \beta_2 \cdot \text{Attendance} + \beta_3 \cdot \text{Parental\_Involvement} + \ldots$$

należy zamienić na wartość numeryczną

Przykład dla 1 wiersza (indeks 0)

$$\hat{y} = \beta_0 + \beta_1 \cdot 23 + \beta_2 \cdot 84 + \beta_3 \cdot 1 + \ldots$$

Zakładając, że Low = 1, Medium = 2, High = 3

Jak ustalić wartość wag β?

- Na początku zainicjować wagi wartościami losowymi (istnieją sposoby bardziej przemyślanej inicjalizacji)
- Rozpocząć proces uczenia maszynowego, którego celem jest otrzymanie/nauczenie się zestawu najlepszych wag.

Więcej na ten temat zostanie powiedziane w przyszłości

Aby móc rozwijać się/uczyć się potrzeba świadomości o popełnianych błędach. W przypadku tematu uczenia maszynowego wprowadza się funkcję straty/błędu oznaczoną najczęściej f_{loss}, f_{cost}

Pozostając przy regresji liniowej:

Suma po wszystkich obserwacjach (punktach)

$$f_{loss}(\hat{y}, y) = \sum \left(\hat{y}(\beta, x^j) - y(x^j)\right)^2$$

wartość rzeczywista

wartość przewidziana przez model

Zwróć uwagę, że zastosowano górny indeks

Rozważmy dane z tabeli zamieszczonej wyżej:

tutaj nie ma exam_score

$$\hat{y}(\beta, x^0) = \beta_0 + \beta_1 \cdot 23 + \beta_2 \cdot 84 + \beta_3 \cdot 1 + \dots$$

predykujemy exam_score dla pierwszej obserwacji

$y(x^0) = 67$
$y(x^1) = 61$

Rzeczywiste wartości exam_score

$$\hat{y}(\beta, x^1) = \beta_0 + \beta_1 \cdot 19 + \beta_2 \cdot 64 + \beta_3 \cdot 1 + \dots$$

Załóżmy, że $\hat{y}(\beta, x^0) = 65$, wtedy błąd dla obserwacji x^0 wynosi

$$\left(\hat{y}(\beta, x^0) - y(x^0)\right)^2 = (65-67)^2 = 4$$

predykcję można uznać za dobrą

Teraz niech $\hat{y}(\beta, x^1) = 52$, wtedy błąd

$$\left(\hat{y}(\beta, x^1) - y(x^1)\right)^2 = (52-61)^2 = 81$$

w tym przypadku predykcja jest już znacznie gorsza

Dlatego błąd zawsze musi uwzględniać wszystkie dane wejściowe/treningowe

Uczenie maszynowe w omawianym przypadku to rozwiązanie zagadnienia

$$\min_{\beta} \left(\sum_j \| \hat{y}(\beta, x^j) - y(x^j) \|_2^2 \right)$$

znaleźć wartość wektora wag β dla którego podane wyrażenie przyjmuje wartość najmniejszą.

Co oznacza symbol $\|\cdot\|$?

$\|\cdot\|$ ← to odwzorowanie nazywane normą (norma)

$$\|\cdot\| : X \to [0, \infty)$$

przestrzeń liniowa (wektorowa) nad ciałem K liczb rzeczywistych lub zespolonych

Norma to uogólnienie pojęcia długości wektora, jednak wystarczy rozumieć, że norma to odległość danego elementu przestrzeni wektorowej od początku tej przestrzeni. Istnieje wiele różnych sposobów mierzenia odległości, czyli istnieje wiele norm, dlatego czasami można spotkać dodatkowe symbole przy $\|\cdot\|$, np. $\|\cdot\|_2$ lub $\|\cdot\|_p$

$$\|a\|_2 = \sqrt{\sum_k a_k^2}\,, \quad \text{np. } \vec{a} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \quad \|\vec{a}\|_2 = \sqrt{3^2 + 2^2 + 1^2} = \sqrt{14}$$

$$\|\vec{a}\|_2^2 = \left(\sqrt{3^2 + 2^2 + 1^2}\right)^2 = 3^2 + 2^2 + 1^2 = 14$$

Sumę $\sum_j \left(\hat{y}(\beta, x^j) - y(x^j)\right)$ można zapisać bardziej zgrabnie, za pomocą notacji macierzowej

Niech
$$X = \begin{bmatrix} 1 & x_1^0 & x_2^0 & x_3^0 & \dots & x_m^0 \\ 1 & x_1^1 & x_2^1 & x_3^1 & \dots & x_m^1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^n & x_2^n & x_3^n & \dots & x_m^n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \quad y = \begin{bmatrix} y(x^0) \\ y(x^1) \\ \vdots \\ y(x^n) \end{bmatrix}$$

Wtedy $(X\beta - y) = e$ ← wektor zawierający wartości błędu na każdej współrzędnej

$$\|e\|_2^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2 + \dots + e_n^2 \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ \vdots \\ e_n \end{bmatrix}$$

Zatem można zapisać:

$$\min_{\beta} \|X\beta - y\|_2^2$$

Przykład dla danych z tabeli ↑

$$X = \begin{array}{c} \begin{array}{ccccc} & \text{Hours\_studied} & \text{Attendance} & \text{Parental\_Involvement} & \text{Access\_to\_Resources} \end{array} \\ \begin{bmatrix} 1 & 23 & 84 & 1 & 3 & \dots \\ 1 & 19 & 64 & 1 & 2 & \dots \\ 1 & 24 & 98 & 2 & 2 & \dots \\ 1 & 29 & 89 & 1 & 2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \end{array}$$

← brak exam_score

exam_score

$$y = \begin{bmatrix} 67 \\ 61 \\ 74 \\ 71 \\ \vdots \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \end{bmatrix}$$

do ustalenia w procesie uczenia maszynowego