Regresja liniowa

Dominik Lau

3 marca 2023

1 Wzór

Regresja liniowa to technika uczenia **z nadzorem**. Mamy dane X, Y.

$$X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ ... & ... \\ x_n & 1 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ ... \\ y_n \end{bmatrix}$$

jedynki w drugiej kolumnie X przydadzą się do skrócenia zapisu. Chcemy wyznaczyć prostą $y = ax_1 + b$, która minimalizuje

$$MSE = \frac{1}{n} \Sigma_{x \in X, y \in Y} (y - ax_0 - b)^2$$

gdzie n to liczba danych a x_0 to pierwsza współrzędna wektora kolumnowego $\vec{x}$ (jednej kolumny macierzy X) więc musimy policzyć **gradient** ale najpierw uproszczenie zapisu:

$$y = ax_0 + b = [a, b] \cdot \begin{bmatrix} x_0 \\ 1 \end{bmatrix} = (\vec{\theta})^T \vec{x}$$

teraz liczymy gradient czyli pochodną po wektorze θ z naszej funkcji błędu

$$\nabla = \frac{\partial}{\partial \theta}((y - \theta x)^2) = -2x(y - \theta x)$$

z ostatniej równości dostajemy wartość **gradientu**, żeby znaleźć minimum (błędu) przyrównujemy do 0

$$-2x(y - \theta x) = 0$$

przechodząc na cały zbiór, czyli na całe macierze (**uwaga!** należy tu brać pod uwagę wymiary wszystkiego, jeszcze je zapiszę poniżej)

$$dimX = n \times 2$$
$$dimY = n \times 1$$
$$dim\theta = 1 \times 2$$

$$-2X^T(Y - X\theta^T) = 0$$
$$X^TY - X^TX\theta^T = 0$$
$$\theta^T = (X^TX)^{-1}X^TY$$

stąd liczymy θ, które minimalizuje nasz błąd i są to współczynniki szukanej prostej

2 Spadek po gradiencie

2.1 Działanie

Czasem nie jesteśmy w stanie w tak łatwy sposób znaleźć wzoru na minima. Wówczas korzystamy z metody spadku po gradiencie. Zapisujemy gradient:

$$\nabla = -2x(y - \theta x)$$

w wersji macierzowej

$$\nabla = -2X^T(Y - X\theta^T)$$

losujemy wartości θ i wykonujemy K (np. 100 iteracji) z każdą iteracją poprawiając θ zgodnie ze wzorem

$$\theta_{i+1} = \theta_i - \kappa\nabla$$

gdzie κ to wspł. kroku, który daje nam większą dokładność parametru (musi być bardzo mały typu 0.001). Iterując w ten sposób przesuwamy się przeciwnie do gradientu co w efekcie zmniejsza wartość funkcji.

2.2 Standaryzacja

Standaryzacja służy do zbliżenia do siebie wartości (tak, żeby $VX = 0$) dzięki czemu metoda spadku po gradiencie działa w mniejszej liczbie iteracji. Żeby ustandaryzować dane trzeba dla X, Y obliczyć wartości X_s, Y_s zgodnie ze wzorem

$$X_s = \frac{X - EX}{\sigma_X}$$

$$Y_s = \frac{Y - EY}{\sigma_Y}$$

E - średnia (**operator!**)
$\sigma = \sqrt{V}$ - odchylenie standardowe
V - wariancja (**operator!**)

3 Polecenia

Dla danych o samochodach (X- masa, Y - spalanie):

1. wyznaczyć θ ze wzoru (3p)
   (a) policzyć ze wzoru θ
   (b) dla wyznaczonego θ policzyć MSE
2. wyznaczyć θ metodą spadku po gradiencie
   (a) standaryzacja danych (1p)
   (b) wyznaczyć θ (1p)