Գծային ռեգրեսիա

Հայկ Կարապետյան

1 Նկարագրություն

Ունենք բնակարանների տվյալների զույգեր $(x_i, y_i)_{i=1}^n, \ x_i \in R^k, \ y_i \in R$: x-ը ներկայացնում է բնակարանի k հատ բնութագրիչները (x_i^1 - սենյակների քանակ, x_i^2 - մակերես, x_i^k - k-րդ բնութագրիչ), իսկ y-ը բնակարանի գինն է: Սահմանենք հետևյալ ֆունկցիան.

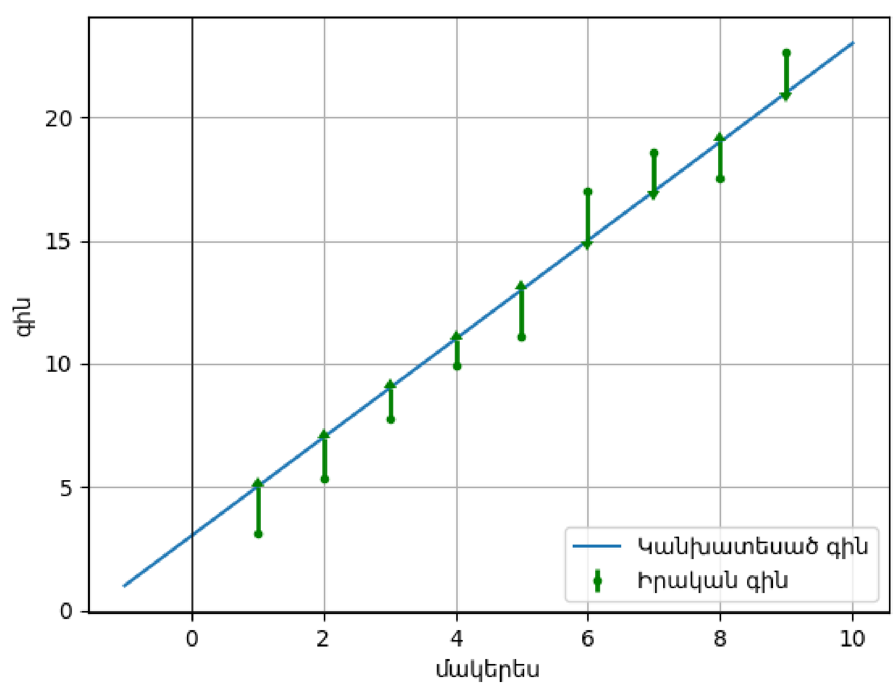
$$f(x) = f(x^1, x^2, ..., x^k) = w^1x^1 + w^2x^2 + ... + w^kx^k + b = w^Tx + b$$

$$x = \begin{bmatrix} x^1 \\ x^2 \\ ... \\ x^k \end{bmatrix}, \ w = \begin{bmatrix} w^1 \\ w^2 \\ ... \\ w^k \end{bmatrix}$$

Մեր նպատակն է գտնել այնպիսի $w^1, w^2, ..., w^k, b$ պարամետրեր, որ $f(x_i) \approx y_i, i = 1, ..., n$ և ոչ միայն մեր ունեցած տվյալների համար: Հետևյալ ֆունկցիայի (գծի) պարամետրերը գտնելը կոչվում է գծային ռեգրեսիա (linear regression): Սահմանենք L^2 հեռավորությունը որպես կորստի ֆունկցիա:

$$L = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2$$

Հետևյալ կորստի ֆունկցիան մինիմիզացնելը կնշանակի, որ մենք հասել ենք մեր նպատակին:



Նկար 1: Կանաչ գիծը ցույց է տալիս կորուստը և մոդելի ուսուցումից հետո հեռավորությունը կանխատեսած ու իրական գնի միջև պետք է լինի նվազագույնը:

Կորուստը նվազագույնի հասցնելու և պարամետրերի արժեքները գտնելու համար ունենք 2 եղանակ.

1. Ածանցյալը հավասարեցնենք զրոյի
2. Օգտագործենք գրադիենտային վայրեջքի մեթոդը

Մեթոդ 1. Վերցնենք $f(x) = X\theta,\ L(\theta) = \frac{1}{n}(X\theta - y)^2$:

$$X = \begin{bmatrix} x_1^1, x_1^2, ..., x_1^k, 1 \\ x_2^1, x_2^2, ..., x_2^k, 1 \\ ... \\ x_n^1, x_n^2, ..., x_n^k, 1 \end{bmatrix},\ \theta = \begin{bmatrix} w^1 \\ w^2 \\ .... \\ w^k \\ b \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \\ .... \\ y_n \end{bmatrix}$$

Լուծումը տեսնելու համար կարող եք միանգամից անցնել (1) հավասարմանը

$$J(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \frac{1}{n}(X\theta - y)^T(X\theta - y)$$

$\frac{1}{n}$ հաստատուն է և փոփոխում է միայն մինիմումի արժեքը, ոչ թե կետը

$$J(\theta) = ((X\theta)^T - y^T)(X\theta - y)$$

$$J(\theta) = (X\theta)^T(X\theta) - (X\theta)^T y - y^T(X\theta) + y^T y$$

Հաշվի առնելով որ $((X\theta)^T y)^T = y^T(X\theta)$

$$J(\theta) = (X\theta)^T(X\theta) - 2y^T(X\theta) + y^T y$$

$$J(\theta) = \theta^T X^T X\theta - 2y^T X\theta + y^T y$$

Հայտնի մատրիցային ածանցյալների բանաձևերն են.

$$\frac{\partial(AX)}{\partial X} = A^T$$

$$\frac{\partial(X^T A)}{\partial X} = A$$

$$\frac{\partial(X^T X)}{\partial X} = 2X$$

$$\frac{\partial(X^T AX)}{\partial X} = AX + A^T X$$

$$\frac{\partial J(\theta)}{\partial \theta} = X^T X\theta + X^T X\theta - 2X^T y = 2X^T X\theta - 2X^T y$$

Գտնելու համար θ-ն, ածանցյալը հավասարեցնենք զրոյի:

$$2X^T X\theta - 2X^T y = 0$$

$$X^T X\theta = X^T y$$

$$\theta = (X^T X)^{-1} X^T y \tag{1}$$

Հետագայում նոր բնակարանի k հատ բնութագրիչ ունենալու դեպքում, կբազմապատկենք θ-ով և կգուշակենք գինը:

Մեթոդ 2. Վերցնենք $f(x) = XW + b,\ L(W,b) = \frac{1}{n}(XW + b - y)^2,\ \alpha = 0.1$:

$$X = \begin{bmatrix} x_1^1, x_1^2, ..., x_1^k \\ x_2^1, x_2^2, ..., x_2^k \\ ... \\ x_n^1, x_n^2, ..., x_n^k \end{bmatrix},\ W = \begin{bmatrix} w^1 \\ w^2 \\ .... \\ w^k \end{bmatrix},\ y = \begin{bmatrix} y_1 \\ y_2 \\ .... \\ y_n \end{bmatrix}$$

$$\nabla L = \left[\frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}\right] = \left[\frac{2X}{n}(XW + b - y), \frac{2}{n}(XW + b - y)\right]$$

Ամեն գրադիենտային վայրեջքի քայլի ընթացքում թարմացնելու ենք կշիռները հետևյալ կերպ.

$$W = W - \alpha\frac{\partial L}{\partial W}$$

$$b = b - \alpha\frac{\partial L}{\partial b}$$

Իսկ ինչպե՞ս կարող ենք ներկայացնել ռեգրեսիան, որպես նեյրոնային ցանց: Ամեն ինչ պարզ է՝ մեկ նեյրոն և ոչ մի ակտիվացիոն ֆունկցիա:

2 Քառակուսային կորուստը ընդդեմ մոդուլով կորստի

Իսկ ինչու ենք օգտագործում քառակուսային կորստի ֆունկցիան, մոդուլի փոխարեն.

$$MSE(Squared) = (f(x) - y)^2 \quad vs \quad MAE(Absolute) = |f(x) - y|$$

$$f(x) = wx + b$$

$$\frac{\partial MSE}{\partial w} = 2(f(x) - y) \times \frac{\partial f(x)}{\partial w} \quad vs \quad \frac{\partial MAE}{\partial w} = signum(f(x) - y) \times \frac{\partial f(x)}{\partial w}$$

Կարող ենք տեսնել, որ MSE կորստի ֆունկցիայի դեպքում ինչքան մոտիկանում ենք y-ին, այնքան մասնակի ածանցյալը փոքրանում է, այսինքն ավելի փոքր քայլերով ենք շարժվում, իսկ MAE-ի դեպքում, անկախ նրանից, թե ինչքան մոտ ենք y-ին, նույն չափով շարժվում ենք դրական կամ բացասական ուղղությամբ՝ կախված signum-ի նշանից: Ավելի փոքր քայլեռով շարժվելը նշանակում է, որ մինչ կշիռի թարմացումը և թարմացումից հետո ստացված կշիռի արժեքների տարբերությունը մեծ չի: Որոշ դեպքերում օգտագործելով MAE կորուստը և ավտոմատ կերպով ուսուցման արագության (learning rate) փոքրացումը կարող են տալ նույն արդյունքը, ինչ MSE կորուստով ուսուցումը: