Կորստի ֆունկցիա: Գրադիենտային վայրեջք

Հայկ Կարապետյան

1 Կորստի ֆունկցիա

Կորստի ֆունկցիան մոդելի ուսուցման ընթացքում նման է ուղեցույցի: Կորստի ֆունկցիան համեմատում է մոդելի գուշակած պիտակը, իրական պիտակի հետ: Նպատակն է նվազագույնի հասցնել կորուստը, որը կստիպի մոդելին կատարել ճիշտ գուշակություններ: Ռեգրեսիոն խնդիրներում կորստի ֆունկցիայի տեսակներից են միջին բացարձակ կորուստը (Mean Absolute Error), միջին քառակուսային կորուստը (Mean Squared Error) և արմատ միջին քառակուսային կորուստը (Root Mean Squared Error): Առաջին ֆունկցիայի դեպքում վերցնում ենք գուշակած արժեքը, հանում իրական արժեքը և վերցնում ստացված արդյունքի մոդուլը:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f(x_i) - y_i^{real}|$$

որտեղ f-ը մոդելի ֆունկցիան է, որը մուտքում
ընդունում է տվյալներ և ելքում տալիս պիտակ:

Երկրորդ ֆունկցիայի դեպքում գուշակած արժեքից հանում ենք իրական արժեքը և բարձրացնում քառակուսի:

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

Օրինակ՝ տան գինը գուշակող մոդել կառուցելիս կորստի ֆունկցիան կարող ենք վերցնել RMSE ($\sqrt{MSE}$):

2 Գրադիենտային վայրեջք

Արդեն ասեցինք, որ մոդելը ուսուցանելիս մեր նպատակն է նվազագույնի հասցնել կորստի ֆունկցիայի արժեքը, այսինքն գտնել ֆունկցիայի մինիմումի կետը: Ամենապարզ դեպքում կարող ենք ֆունկցիան ածանցել և հավասարեցնել զրոյի: Մի քանի փոփոխականի դեպքում մեզ հարկավոր կլինի կազմել համակարգ տարբեր $x_1, x_2, ..., x_n$ արժեքներ փորձարկելով: Բայց հնարավոր է, որ ֆունկցիան շատ բարդ լինի և համակարգը լուծում չունենա ու մենք հստակ չկարողանանք գտնել մինիմումի կետը: Այդ դեպքում, ինչպե՞ս կարող ենք գտնել կամ մոտարկել ֆունկցիայի մինիմումի կետը: Մաթեմատիկայում կա գրադիենտ վեկտոր հասկացություն: Գրադիենտ վեկտորը ֆունկցիայի մասնակի ածանցյալներից կազմված վեկտորն է: Այն ցույց է տալիս ֆունկցիայի ամենաարագ աճման ուղղությունը: Եթե ցանկանում ենք գտնել ֆունկցիայի մինիմումը, մեզ պետք է շարժվել ամենաարագ նվազման ուղղությամբ, այսինքն գրադիենտ վեկտորին հակառակ: Եկեք նայենք մեկ օրինակ:
Մեր կորստի ֆունկցիան ունի հետևյալ տեսքը՝ $L = (w-3)^2$: Եթե ֆունկցիան ածանցենք և հավասարեցնենք զրոյի, կտեսնենք որ նրա մինիմումի կետն է $w = 3$:

$$L' = 0$$
$$2*(w-3) = 0$$
$$w = 3$$

Եկեք մինիմումի կետը գտնենք շարժվելով գրադիենտի հակառակ ուղղությամբ: Մեթոդը կոչվում է գրադիենտային վայրեջք (gradient descent): Գրադիենտ վեկտորը նշանակենք ∇-ով:

$$\nabla L = \left[\frac{\partial L}{\partial w}\right] = [2*(w-3)]$$

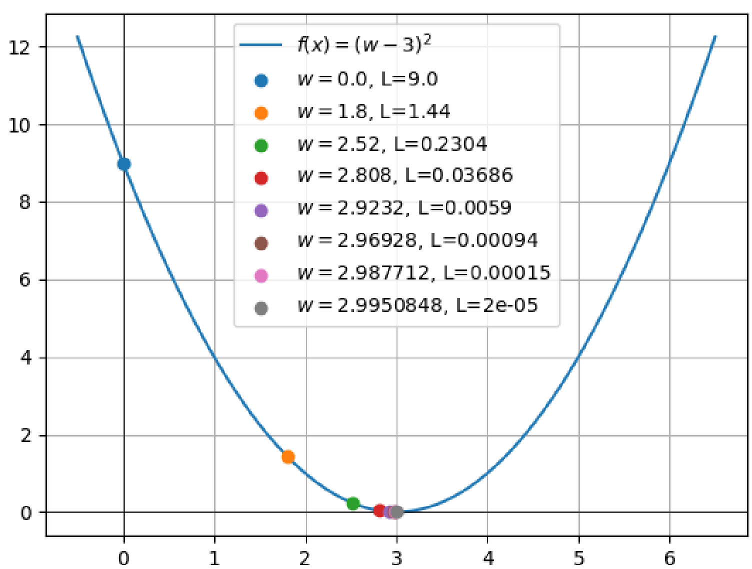
Վերցնենք w-ի սկզբնարժեքը 0: Գրադիենտի հակառակ ուղղությամբ շարժվելու համար w-ից հանում ենք գրադիենտ վեկտորը w կետում: α-ն ցույց է տալիս, թե ինչքան է պետք շարժվել գրադիենտի հակառակ ուղղությամբ: α-ն կոչվում է ուսուցման արագություն (learning rate): Հետևյալ օրինակում վերցնենք $\alpha = 0.3$:

$$w = w + \alpha\left(-\frac{\partial L}{\partial w}\right) = w - \alpha\frac{\partial L}{\partial w} = 0 - 0.3 * 2 * (0 - 3) = 1.8$$

Մեկ քայլ կատարելուց հետո կարող ենք տեսնել, որ դեռ չենք հասել մինիմումի կետ: Հետևաբար այս գործողությունը պետք է կատարել այնքան ժամանակ մինչև $w_{\text{նոր}}$ և $w_{\text{հին}}$ արժեքների տարբերությունը փոքր լինի օրինակ 10^{-2}-ից: Շարունակենք.

$$w = w - \alpha\frac{\partial L}{\partial w} = 1.8 - 0.3 * 2 * (1.8 - 3) = 2.52$$
$$w = 2.52 - 0.3 * 2 * (2.52 - 3) = 2.808$$
$$w = 2.808 - 0.3 * 2 * (2.808 - 3) = 2.9232$$
$$w = 2.9232 - 0.3 * 2 * (2.9232 - 3) = 2.96928$$
$$w = 2.96928 - 0.3 * 2 * (2.96928 - 3) = 2.987712$$
$$w = 2.987712 - 0.3 * 2 * (2.987712 - 3) = 2.9950848$$
$$\text{տարբերություն} = 2.9950848 - 2.987712 = 0.0073728 < 10^{-2}$$

Ստացանք որ ֆունկցիան ընդունում է իր փոքրագույն արժեքը, երբ $w = 2.9950848$ (Նկար 1):



Նկար 1: w պարամետրի և կորստի արժեքները՝ ամեն գրադիենտային վայրէջքի քայլից հետո

Ավելի լավ մոտարկում ստանալու համար կարող ենք 10^{-2} դարձնել 10^{-5}: α-ն արժեք է, որը ընտրում ենք մենք, այլ ոչ թե փոփոխվում է ուսուցման ընթացքում:
Երկու փոփոխական ունենալու դեպքում, մեկ քայլի ընթացքում նաև կփոփոխեինք երկրորդի արժեքը:

$$w_1^{\text{նոր}} = w_1^{\text{հին}} - \alpha\nabla L(w_1^{\text{հին}}, w_2^{\text{հին}})$$
$$w_2^{\text{նոր}} = w_2^{\text{հին}} - \alpha\nabla L(w_1^{\text{հին}}, w_2^{\text{հին}})$$

Եկեք նայենք ևս մեկ օրինակ, երբ ունենք տվյալների զույգեր $(x_i, y_i),\ i = 1, 2, 3$: Վերցնենք 2x+3 ֆունկցիան տվյալների զույգեր ստանալու համար և փորձենք գրադիենտային վայրեջքի միջոցով գտնել այդ ֆունկցիան:

id	x	y
1	1	5
2	4	11
3	8	19

Աղյուսակ 1: x, y տվյալների զույգեր

Կառուցենք մեկ նեյրոնից բաղկացած մոդել և փորձենք մոտարկել այդ ֆունկցիան: Կորստի ֆունկցիան վերցնենք MSE:

$$f_{model}(x) = wx + b$$

Սկզբնարժեքավորենք w=1 և b=0: $\alpha = 0.1$ Գրենք գրադիենտ վեկտորը.

$$L = (wx + b - y_{real})^2$$

$$\nabla L = \left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right] = [2 * (wx + b - y_{real}) * x, 2 * (wx + b - y_{real})]$$

Թարմացնենք ուսուցանվող պարամետրերը կատարելով գրադիենտային իջեցում՝ օգտագործելով մեր բոլոր տվյաները (միջինացնելով գրադիենտ վեկտորները):

$$\frac{\partial L}{\partial w}(x_1) = 2 * (1 * 1 + 0 - 5) = -8$$

$$\frac{\partial L}{\partial w}(x_2) = 2 * (1 * 4 + 0 - 11) = -14$$

$$\frac{\partial L}{\partial w}(x_3) = 2 * (1 * 8 + 0 - 19) = -22$$

$$\frac{\partial L}{\partial w} = \frac{1}{3}\left(\frac{\partial L}{\partial w}(x_1) + \frac{\partial L}{\partial w}(x_2) + \frac{\partial L}{\partial w}(x_3)\right) = -\frac{44}{3}$$

$$w = w - \alpha\frac{\partial L}{\partial w} = 1 + 0.1 * \frac{44}{3} = \frac{37}{15}$$

Նույն գործողությունը կատարում ենք b պարամետրի արժեքը թարմացնելու համար և կրկնում ենք այնքան ժամանակ, մինչև տարբերությունը առկա արժեքի և թարմացված արժեքի միջև փոքր լինի 10^{-5}-ից կամ մոդելի ճշգրտությունը բարձր լինի սահմանված արժեքից: