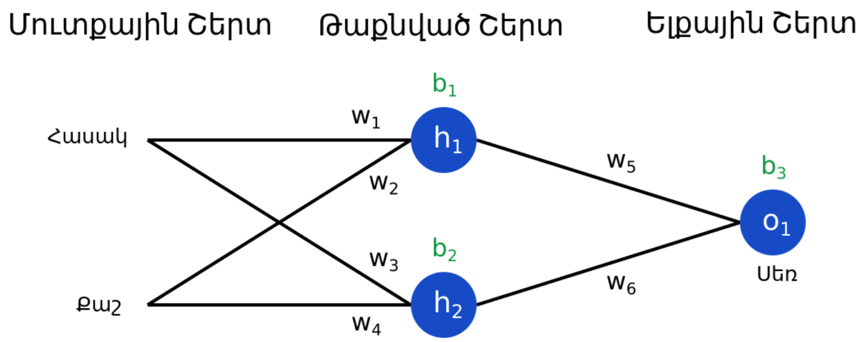
Կշիռների սկզբնարժեքավորում

Հայկ Կարապետյան

1 Միանման սկզբնարժեքավորում

Կարո՞ղ ենք գծային կամ լոգիստիկ ռեգրեսիայի պարամետրերը սկզբնարժեքավորենք 0-ով: Պատասխանն է՝ այո: Իսկ կարո՞ղ ենք կամայական նեյրոնային ցանցի բոլոր պարամետրերը սկզբնարժեքավորենք 0-ով: Պատասխանն է՝ ոչ: Դիտարկենք նեյրոնային ցանցի օրինակ, բաղկացած երկու նեյրոնով մեկ թաքնված շերտից և մեկ ելքային նեյրոնից (Նկար 1): Մուտքային արժեքներն են մարդու քաշը, հասակը և պետք է գուշակել մարդու սեռը:



Նկար 1: Մեկ թաքնված շերտից բաղկացած նեյրոնային ցանց

$$h_1 = f(w_1x_1 + w_2x_2 + b_1), \quad h_2 = f(w_3x_1 + w_4x_2 + b_2)$$
$$o_1 = g(w_5h_1 + w_6h_2 + b_3)$$
$$L(w) = L(w_1, \dots, w_6) = (o_1 - y)^2$$

Սկզբնարժեքավորենք կշիռները հետևյալ կերպ

$$w_1^0 = w_3^0 = c_1,\ w_2^0 = w_4^0 = c_2,\ w_5^0 = w_6^0 = c_3,\ b_1^0 = b_2^0 = c_4$$
$$w_i^1 = w_i^0 - \alpha\frac{\partial L}{\partial w_i}, i = 1, \dots, 6$$

Գրենք մասնակի ածանցյալները w_1-ից w_4 կշիռների համար

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial o_1} \times g'(w_5h_1 + w_6h_2 + b_3) \times w_5 \times f'(w_1x_1 + w_2x_2 + b_1) \times x_1$$
$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial o_1} \times g'(w_5h_1 + w_6h_2 + b_3) \times w_5 \times f'(w_1x_1 + w_2x_2 + b_1) \times x_2$$
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial o_1} \times g'(w_5h_1 + w_6h_2 + b_3) \times w_6 \times f'(w_3x_1 + w_4x_2 + b_1) \times x_1$$
$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial o_1} \times g'(w_5h_1 + w_6h_2 + b_3) \times w_6 \times f'(w_3x_1 + w_4x_2 + b_1) \times x_2$$

Կարող ենք տեսնել, որ $\dfrac{\partial L}{\partial w_1} = \dfrac{\partial L}{\partial w_3}, \quad \dfrac{\partial L}{\partial w_2} = \dfrac{\partial L}{\partial w_4}$

Սա նշանակում է որ w_1 և w_3 կշիռները նույն կերպ են թարմացվելու, այսինքն անիմաստ է ունենալ երկու առանձին նեյրոն մեկ է նրանք նույն բանն են սովորելու: Սա է խնդիրը, որ նեյրոնային ցանցերում բոլոր պարամետրերը չեն կարող միանման սկզբնարժեքավորվել կամ զրոներով սկզբնարժեքավորվել: Իսկ գծային կամ լոգիստիկ ռեգրեսիայի դեպքում մենք ունենք միայն մեկ նեյրոն և կարող ենք այդ նեյրոնի պարամետրերը զրոյով սկզբնարժեքվորել:

2 Գլորոտ-Խավիեր Սկզբնարժեքավորում

Սկզբնարժեքավորումը արվում է միայն մեկ անգամ, մոդելի ուսուցումից առաջ: Պատահական կերպով սկզբնարժեքավորելիս հնարավոր է կշիռները շատ փոքր լինեն և մեծ ցանցերի դեպքում մուտքային տվյալները զրոյանան մինչև վերջ հասնելը: Կամ կարող է լինել հակառակը՝ կշիռները մե արժեքներով սկզբնարժեքավորվեն և մինչև վերջ հասնելը մուտքային տվյալները չափից շատ մեծանան: Գլորոտ-Խավիեր սկզբնարժեքավորումը առաջարկում է պահել նույնը մուտքային տվյալների միջակայքը, որպեսզի նույնիսկ խորը ցանցերում մինչև վերջ հասնելը զրոյացած չլինի մուտքը: Դա առաջարկվում է կատարել հետևյալ կերպ (մաթեմատիկայի մեջ չխորանալու համար անցիր (1) բանաձևին):
Ունենք մուտքային X, W պարամետրեր և Y ելք:

$$Y = W_1X_1 + W_2X_2 + \cdots + W_nX_n$$

Ենթադրենք որ$\{X_i\}_{i=1}^n$ և $\{W_i\}_{i=1}^n$անկախ պատահական փոփոխականներ են, որոնք միարժեք նույն ձև են բաշխված

$$E(X_1) = E(X_i) = E(W_1) = E(W_i) = 0, \quad i = 1, \dots n$$

Նկատենք որ

$$Var(W_i, X_i) = E(X_i)^2 Var(W_i) + E(W_i)^2 Var(X_i) + Var(W_i)Var(X_i) =$$
$$= Var(W_i)Var(X_i) = Var(W_1)Var(X_1)$$
$$Var(Y) = Var(W_1X_1 + W_2X_2 + \cdots + W_nX_n) = nVar(W_1)Var(X_1)$$

Եթե ցանկանում ենք մուտքային և ելքային արժեքները ունենան նույն բաշխումը, մեզ հարկավոր է $Var(Y) = Var(X_1)$ և այդտեղից կստանանք որ

$$Var(W_1) = \frac{1}{n}$$

Որտեղ n-ը շերտի նեյրոնների քանակն է
Հետադարձ տարածման ժամանակ գրադիենտների բաշխումը փոխվում է
Դա նույնպես հաշվի առնելու համար կատարենք հետևյալ փոփոխությունը

$$Var(W_1) = \frac{2}{n_{input} + n_{output}} \tag{1}$$

n_{input} – նախորդ շերտի նեյրոնների քանակ, $\quad n_{output}$ – տվյալ շերտի նեյրոնների քանակ

Կող գրելիս, եթե ցանկանանք հետևյալ վարիանսը ապահովել մեզ հարկավոր է կշիռները սկզբնարժեքավորել հետևյալ յունիֆորմ բաշխումից, որի սահմանները գրված են փակագծերում:

$$WU\left[-\frac{\sqrt{6}}{\sqrt{n_{input} + n_{ouput}}}, \ \frac{\sqrt{6}}{\sqrt{n_{input} + n_{ouput}}}\right]$$