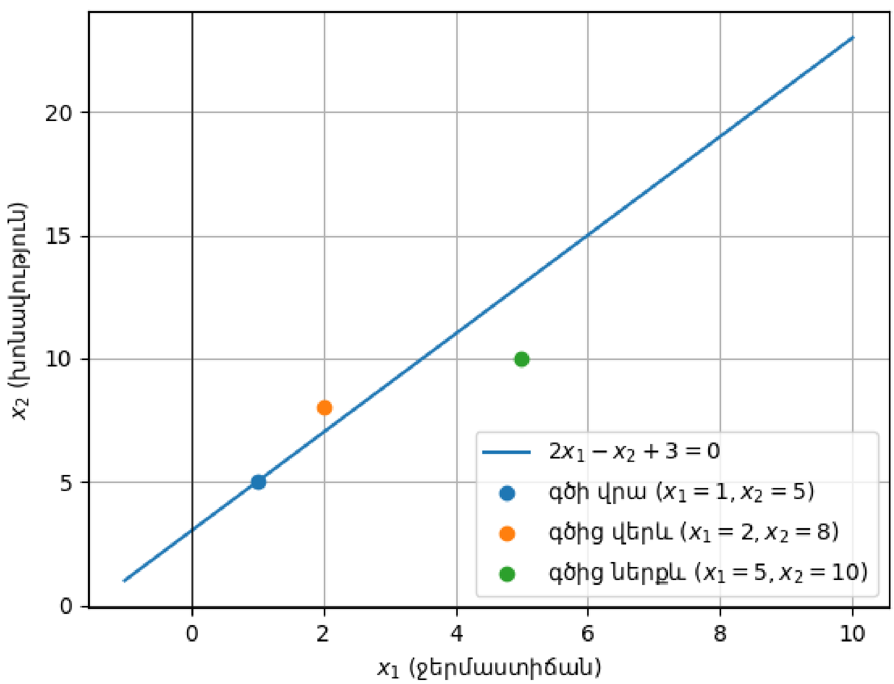
# Լոգիստիկ ռեգրեսիա

Հայկ Կարապետյան

Ունենք օրվա տարբեր պահերի եղանակի տվյալների զույգեր $(x_i, y_i)_{i=1}^{n},\ x_i \in R^k,\ y_i \in \{0, 1\}$: x-ը ներկայացնում է եղանակի $k$ հատ բնութագրիչները (ջերմաստիճան, խոնավություն), իսկ y-ը եղանակի պիտակն է (անձրև կգա՝ 0, անձրև չի գա՝ 1): Մեզ հարկավոր է ֆունկցիա, որին մուտքում կկարողանանք տալ նկար և այն կասի՝ պատկերված է շուն, թե կատու: Գծային ռեգրեսիայից գիտենք, որ գծի ֆունկցիա կարելի է ստանալ օգտագործելով մեկ նեյրոնից բաղկացած նեյրոնային ցանց: Հայտնի է գծի ֆունկցիայի ընդհանուր հավասարումը.

$$w_1x_1 + w_2x_2 + b = 0$$

$w_1, w_2$ և $b$ արժեքները որոշվում են ուսուցման ընթացքում, իսկ $x_1$ և $x_2$ մուտքային տվյալներն են: Եթե $x_1$ և $x_2$ արժեքները տեղադրելով հավասարման մեջ ստանում ենք 0, նշանակում կետը գտնվում է գծի վրա: Եթե մեծ է զրոյից ապա գտնվում է գծից ներքև, եթե փոքր է զրոյից, ապա գտնվում է գծից վերև: Մեզ հարկավոր է ֆունկցիա, որը կվերադարձնի 0, երբ արժեքը լինի




Նկար 1: Գծի միջոցով հնարավոր է կատարել տվյալների դասակարգում

բացասական և 1 հակառակ դեպքում: Այդպիսի ֆունկցիայի օրինակ է.

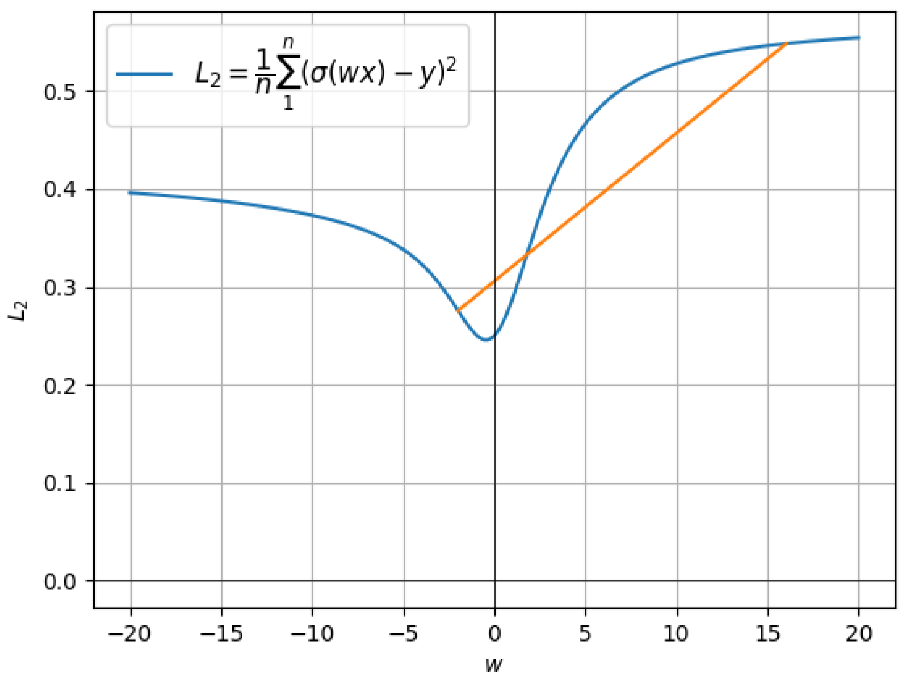
$$f(x_1, x_2) = \frac{1 + signum(w_1x_1 + w_2x_2 + b)}{2}$$

Արդյո՞ք հնարավոր է կառուցել նեյրոնային ցանց և ուսուցանել այն հետևյալ ֆունկցիայի հիման վրա: Նեյրոնայի ցանցը կունենա մեկ նեյրոն և ակտիվացիոն ֆունկցիան կվերցնենք

$\varphi(x) = \dfrac{1 + signum(x)}{2}$: Կորստի ֆունկցիան վերցնենք քառակուսային կորուստը:

$$L = (f(x_1, x_2) - y)^2$$
$$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1} = w_1 - 2\alpha(f(x_1, x_2) - y)\frac{\partial f(x_1, x_2)}{\partial w_1}$$
$$\frac{\partial f(x_1, x_2)}{\partial w_1} = \frac{\partial signum(w_1x_1 + w_2x_2 + b)}{\partial w_1} = x_1 \times 0 = 0$$

Գրադիենտային վայրեջք կիրառելիս կարող ենք տեսնել, որ ակտիվացիոն ֆունկցիայի ածանցյալը հավասար է 0-ի: Այսինքն նեյրոնային ցանցը ոչինչ չի սովորի: Այդ պատճառով պետք է փոխարինել ակտիվացիոն ֆունկցիան և այս դեպքում կարող ենք կիրառել sigmoid ակտիվացիոն ֆունկցիան: Ինչպես գիտենք sigmoid-ի ածանցյալը կարող ենք արտահայտել հենց իրենով և այս դեպքում w-ի արժեքը կարող ենք թարմացնել:
Քառակուսային կորուստը օգտագործելու դեպքում կարող ենք տեսնել, որ կորստի ֆունկցիան չի լինում ուռուցիկ (Նկար 2): Այդ պատճառով մեզ հարկավոր է կորստի ֆունկցիա, որը կլինի




Նկար 2: Քառակուսային կորուստը օգտագործելու դեպքում ֆունկցիան ուռուցիկ չի լինում (երկու կետերը իրար միացնելիս հատում ենք գրաֆիկը)

ուռուցիկ sigmoid ակտիվացիոն ֆունկցիան օգտագործելիս: Կորստի ֆունցկիան, որը օգտագործում են լոգիստիկ ռեգրեսիայում (logistic regression) կոչվում է cross-entropy: Այժմ դիտարկենք նրա առաջացման ուղիներից մեկը:
Վերցնենք հետևյալ տվյալների զույգը $(x_i, 1),\ x_i \in R^k,\ y_i = 1$.

$$f(x) = \sigma(wx + b)$$
$$P(label = 1|x_i) = f(x_i)$$
$$P(label = 0|x_i) = 1 - f(x_i)$$

Մեզ հարկավոր է մաքսիմիզացնել $f(x_i)$, երբ $y_i = 1$ և պետք է մաքսիմիզացնել $1 - f(x_i)$, երբ

$y_1 = 0$.

$$P(label = y_i|x_i) = \begin{bmatrix} f(x_i), \text{երբ } y_1 = 1 \\ 1 - f(x_i), \text{երբ } y_i = 0 \end{bmatrix} = f(x_i)^{y_i}(1 - f(x_i))^{1-y_i}$$

Սահմանենք հետևյալ կորստի ֆունկցիան, որը հարկավոր է մաքսիմիզացնել.

$$L = f(x_i)^{y_i}(1 - f(x_i))^{1-y_i}$$

Խորը ուսուցման մեջ ավելի ընդունված է մինիմիզանցել կորստի ֆունկցիան, այդ պատճառով բազմապատկենք -1-ով

$$-L = -f(x_i)^{y_i}(1 - f(x_i))^{1-y_i}$$

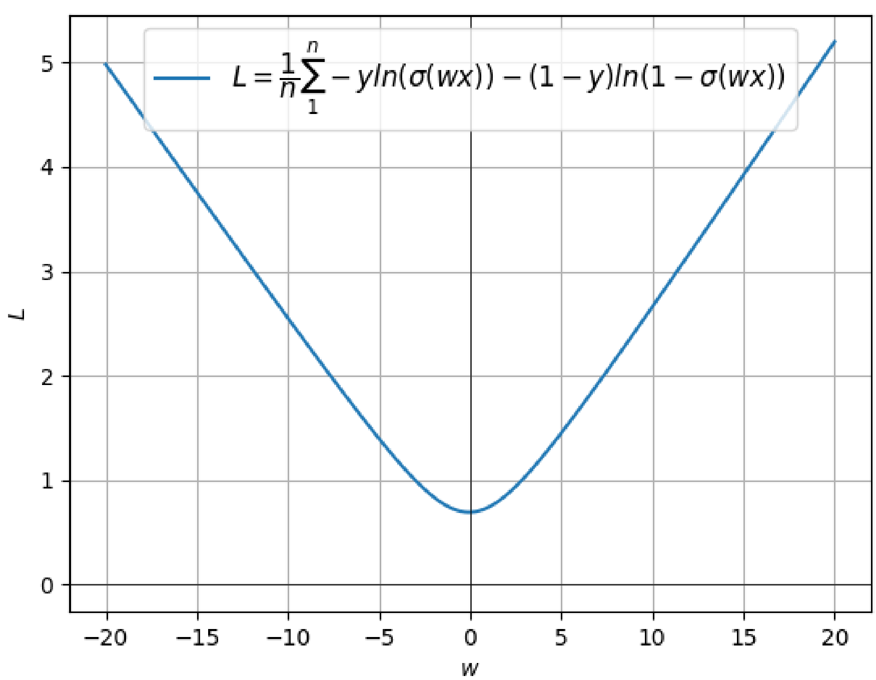
Ավելի գեղեցիկ տեսք ստանալու համար լոգարիթմենք հետևյալ հավաարումը

$$-ln(L) = -y_i ln(f(x_i)) - (1 - y_i)ln(1 - f(x_i))$$

Լոգարիթմը և մինուս նշանը միայն փոխում են կորստի ֆունկցիայի արժեքը, ոչ թե մինիմումի կետը (w, b)

$$L = -y_i ln(f(x_i)) - (1 - y_i)ln(1 - f(x_i))$$

Ստացված կորստի ֆունկցիան կոչվում է cross-entropy կորստի ֆունկցիա: Այս ֆունկցիան ուռուցիկ է (Նկար 3) և մինիմզացնելով այն, կարող ենք կատարել կլասիֆիկացիա: Օրինակ՝ մուտքում տալով ջերմաստիճանը և խոնավությունը, արդյունքում ստանալ ինչքան հավանականությամբ անձրև կգա:




Նկար 3: cross-entropy կորուստը օգտագործելու դեպքում ֆունկցիան լինում է ուռուցիկ (Երկու նկարներում էլ չենք օգտագործում բիաս՝ եռաչափ տարածությունից խուսափելու նպատակով:)

Եկեք դիտարկենք մեկ խնդրի օրինակ.

| $id$ | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| 1 | 15 | 0.5 | 0 |
| 2 | 25 | 0.8 | 1 |
| 3 | 22 | 0.7 | 1 |

Աղյուսակ 1: $x_1$−ջերմաստիճան, $x_2$−խոնավություն, $y = 1$ անձրև կգա, $y = 0$ անձրև չի գա

$$f(x_1, x_2) = \sigma(w_1 x_1 + w_2 x_2 + b)$$

$$L = \frac{1}{3}\sum_{i=1}^{3} -y^i ln(f(x_1^i, x_2^i)) - (1 - y^i) ln(1 - f(x_1^i, x_2^i))$$

$$\frac{\partial L}{\partial w_1} = -y^i \frac{1}{f(x_1^i, x_2^i)} \times \frac{\partial f(x_1^i, x_2^i)}{\partial w_1} - (1 - y^i)\frac{1}{1 - f(x_1^i, x_2^i)} \times \left(-\frac{\partial f(x_1^i, x_2^i)}{\partial w_1}\right)$$

$$\frac{\partial f(x_1, x_2)}{\partial w_1} = \sigma(w_1 x_1 + w_2 x_2 + b)(1 - \sigma(w_1 x_1 + w_2 x_2 + b))x_1$$

$$w_1 = 0.2, w_2 = -6, b = 0, \alpha = 0.1$$

$$x_1^1 = 15, x_2^1 = 0.5, y^1 = 0$$

Անհրաժեշտ փոփոխականները և արժեքները սահմանելուց հետո կարող ենք անցնել գրադիենտային վայրեջքի քայլին

$$\frac{\partial L}{\partial w_1}(x_1^1, x_2^1, y^1) = -y^1 \frac{1}{f(x_1^1, x_2^1)} \times \frac{\partial f(x_1^1, x_2^1)}{\partial w_1} - (1 - y^1)\frac{1}{1 - f(x_1^1, x_2^1)} \times \left(-\frac{\partial f(x_1^1, x_2^1)}{\partial w_1}\right) =$$

$$= 0 - 1 \times \frac{1}{1 - 0.5} \times (-0.5 \times 0.5 \times 15) = 7.5$$

Հետևյալ քայլը կատարում ենք նաև $(x_1^2, x_2^2, y^2)$ և $(x_1^3, x_2^3, y^3)$ զույգերի համար

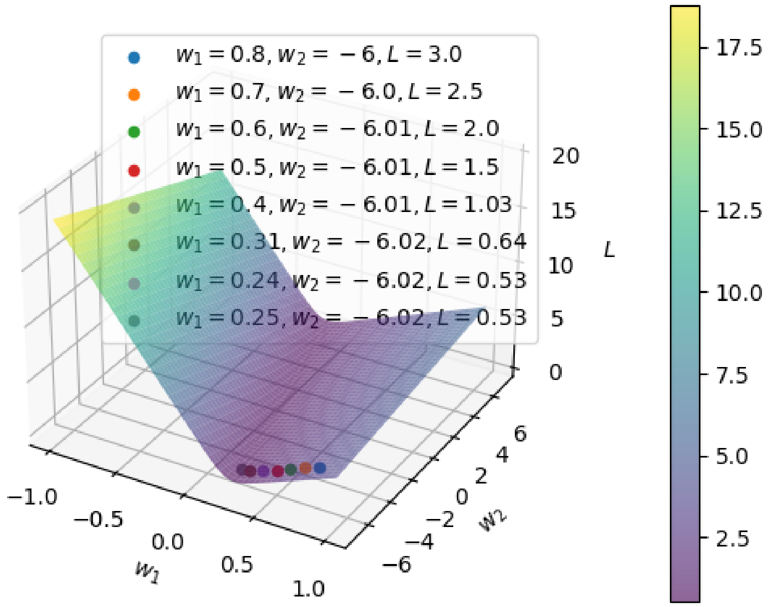
$$w_1 = w_1 - \alpha \times \frac{1}{3}\sum_{i=1}^{3} \frac{\partial L}{\partial w_1}(x_1^i, x_2^i, y^i)$$

Թարմացման հետևյալ եղանակը կիրառում ենք նաև $w_2$ և $b$ պարամետրերի համար

Փորձենք պատկերել կորստի ֆունկցիայի գրաֆիկը կախված $w_1$ և $w_2$ պարամետրերից: Սկզբնարժեքավորենք պարամետրերը հետևյալ կերպ.

$$w_1 = 0.8,\ w_2 = -6,\ b = 0$$

Նկար 4-ում պատկերված է կորստի և պարամետրերի արժեքները ամեն գրադիենտային վայրեջքի քայլից հետո: Օրինակում բիասի առանցքը պատկերված չէ և արժեքը չի թարմացվում, այդ պատճառով կորուստը շատ չի մոտենում 0-ի, բայց մոտենում է իր մինիմում արժեքին: Լոգիստիկ ռեգրեսիան նեյրոնային ցանցի տեսքով ներկայացնելիս ունենալու ենք մեկ նեյրոն և ակտիվացիոն ֆունկցիան sigmoid:
Իսկ կարո՞ղ ենք կիրառել լոգիստիկ ռեգրեսիա, երբ պիտակների քանակը մեծ է 2-ից: Պատասխանն է այո: 3 պիտակից բաղկացած կլասիֆիկացիայի դեպքում կարող ենք վերջին շերտում վերցնել 3 նեյրոն և նրանցից ամեն մեկը կասի՝ մուտքային տվյալը ինչքան հավանականությամբ է պատկանում տվյալ պիտակին: Օրինակ՝ շուն, կատու, փիղ կլասիֆիկացիայի դեպքում կարող ենք վերցնել 3 նեյրոն և առաջին նեյրոնը կասի շուն է, թե շուն չէ, երկրորդը կասի կատու է, թե կատու չէ, երրորդը կասի փիղ է, թե փիղ չէ: Հետևյալ արդյունքներից ելնելով մենք կկարողանանք ասել մուտքային նկարը պատկանում է մեր 3 պիտակներից ինչ որ մեկին (շուն, կատու, փիղ), թե ոչ մեկին էլ չի պատկանում: Ոչ մեկին չպատկանելու դեպքը շատ կարևոր է, քանի որ մեզ պետք չէ, որ մոդելը մարդու նկար տեսնելիս նրան տա մեր 3 պիտակներից ինչ որ մեկը: Հետևյալ մեթոդը կոչվում է բոլորը մեկի դեմ (one vs all):




Նկար 4: cross-entropy կորստի ֆունկցիան կախված $w_1$ և $w_2$ պարամետրերի արժեքներից: Գրադիենտային վայրեջք և պարամետրերի թարմացում: