ԽՈՐԸ (ՄԵՔԵՆԱՅԱՅՎԱԾ) ՈՒՍՈՒՑՈՒՄ

Առնակ Պապայան
16/08/2021.

## 0) Նախաբան

Քանի որ իմ սպասածից շատ ավելի շատ գրանցվողներ են եղել այս դասընթացին՝ ինձ համար կարևոր է հասկանալ ներկաների գիտելիքների մակարդակը: Այդ նպատակով ձեզ կուղղեմ 3 հարց: Խնդրում եմ գրելով չպատասխանել: Ձեռք բարձրացրեք՝ եւ ձեզ ձայն կտամ:

Հ.1 Ի՞նչ է նեյրոնային ցանցը: Ընդունված են միայն մեկ բառանոց պատասխաններ:

Պ.1 Սրբապատկերում:

Հ.2. Ի՞նչ տարբերություն կա խորը ուսուցման և մեքենայական ուսուցման միջև: Կարո՞ղ եք ասել, որ մեկը մյուսի մասն է կազմում:

Պ.2 Խորը ուսուցումը կազմում է մեքենայական ուսուցման մասը: Այն ներառում է միայն խորը նեյրոնային ցանցեր օգտագործող ալգորիթմները:

Հ.3. Ո՞րն է ավելի լավ ալգորիթմ՝ բինար դասակարգման համար, լոգիստիկ ռեգրեսիա՞ն թե՞ ոչ մի թաքնված շերտ չունեցող նեյրոնային ցանցը:

Պ.3. Դրանք նույն բանն են:

Ինչ չի լինի այս դասընթացում՝

1) Շների ու կատուների նկարներ
2) (Պայծառակերպ) մաթ. ապարատի սահմանումը:
3) Անալիզի, հանրահաշվի ու գրաֆների սահմանումը:
4) 20 րոպեից ավել տևող ապացույց

Ինչ կլինի այս դասընթացում՝

1) Խորը ուսուցմանը վերաբերվող մաթ. գաղափարների ֆորմալ սահմանում:
2) Բաց խնդիրների ֆորմալ և ոչ ֆորմալ ձևակերպում:
3) Որոշ մաթեմատիկա՝ առանց թվերի, միայն պատկերով:
4) 20 րոպեից պակաս տևող ապացույցներ:

---

* Բաց նկարազարդումները՝ կարող են լինել փոքրիկ հանձնարարություններ:

* Խնդրում եմ ընդհանուր բնույթի հարցերը ուղղել դասից հետո: Պատճառ ժամանակի՝ միայն կընդունեմ հասկանալի-չհասկանալիությանը վերաբերվող հարցեր:

* Շատ հավանական է, որ նախապատրաստված ծրագիրը ամբողջությամբ չկարողանանք ծածկել:

## 1) Կարևոր նշանակումներ և սահմանումներ

Այս դասընթացում կքննենք միայն վերահսկվող ուսուցման ձևը (supervised learning): Խնդիրը ձևակերպված է հետևյալ կերպ: Ունենք նույն կատեգորիային պատկանող $n$ օբյեկտ: Օրինակ՝ լատինական այբուբենով գրված $n$ հատ պատմվածք: ($n$-ը ոչ շատ փոքր բնական թիվ է, $n \in \mathbb{N}$:) Այդ $n$ օբյեկտներից յուրաքանչյուրի վրա կատարված է $d$ չափում: Ուստի օբյեկտը ներկայացված է ստացված $d$-չափանի վեկտորի տեսքով: Այսպես ստանում ենք

$$X_1, X_2, \dots, X_n \in \mathbb{R}^d$$

վեկտորների հաջորդականությունը: Պատմվածքների օրինակով՝ չափումները կարող են լինել բառերի միջին երկարությունը, բառի մեջ ձայնավորների միջին համամասնությունը, մի նախադասության մեջ բառերի միջին քանակը, և այլն:

Այս $n$ օբյեկտները պիտակավորվել են վերահսկիչի կողմից: $i$-րդ օբյեկտը ստացել է $Y_i$ պիտակը: Այսպիսով մեր ձեռքի տակ ունենք

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

հաջորդականությունը, որը կանվանենք ընտրանք (sample): Նպատակն է ընտրանքի հիման վրա գտնել մի կանոն, որը կպիտակավորի կամայական նոր օբյեկտ՝ առանց վերահսկիչի օգնության:

## Ենթադրություն

Գոյություն ունի $P^*$ հավանականային չափ՝ որոշված $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}$ բազմության վրա, այնպիսին որ $(X_1, Y_1), \dots, (X_n, Y_n)$-ը անկախ պատահական վեկտորներ են $P^*$ բաշխմամբ: Համառոտագրելով,

$$\Pr\big((X_i, Y_i) \in A\big) = P^*(A) \quad \forall i \in \{1, \dots, n\}$$

կամայական չափելի $A$ բազմության համար:

Պատկերների օրինակում՝ $Y_i$-ն կարող է լինել $i$-րդ պատկերի լեյբլը (անգղերը, ֆլամինգոները, զեբրաները, և այլն): Այդ դեպքում լեյբլները նախապես համարակալված են՝

1 = անգղ. 2 = ֆլամ. 3 = զեբր., ...

և ենթադրում, որ $\mathcal{Y} = \{1, 2, \dots, K\}$, որտեղ $K$-ն բոլոր հնարավոր լեյբլների քանակն է:

---

Այսպիսով, մեքենայական ուսուցման խնդիրն է գտնել արտապատկերում՝

$$f : \mathcal{X} \to \mathcal{Y}$$

այնպիսին, որ $f(X)$-ը մոտ է $Y$-ին ոչ միայն արդեն իսկ դիտարկված $(X, Y)$ զույգերի համար, այլ $P^*$ բաշխումից գեներացված զույգերի $\underbrace{\text{մեծ մասի}}$ համար:

Նկատենք, որ բաշխումը կարևոր է հենց այդ "մեծ մասը" չափել կարողանալու համար: 4

<u>Սահմ.</u> $f : \mathcal{X} \to \mathcal{Y}$ չափելի ֆունկցիան կանվանենք կանխատեսիչ։ ▨

$f$ կանխատեսիչը լավն է, եթե

$$P^*\left(\{(x,y) : f(x) \neq y\}\right) \qquad (*)$$

հավանականությունը փոքր է։

Հիշեցնենք, որ $P^*$ հավանականային չափը մեզ հայտնի չէ։ Հետևաբար, պրակտիկայում մենք չենք կարող հաշվել $(*)$ հավանականությունը։

## 2) Նեյրոնային ցանց

Փնտրել ամենալավ կանխատեսիչը՝ բոլոր հնարավոր կանխատեսիչների մեջ, կլիներ շատ բարդ և ոչ այնքան խելամիտ մոտեցում։ Փոխարենը՝ կարելի է նախապես առանձնացնել կանխատեսիչների մի դաս, որի տարրերը ունեն որոշակի հատկություններ, և փնտրել լավագույն կանխատեսիչն այդ դասի մեջ։ Այդպիսի դասի օրինակ են նեյրոնային ցանցերը (ՆՑ)։

* 1 թաքնված շերտով ՆՑ՝

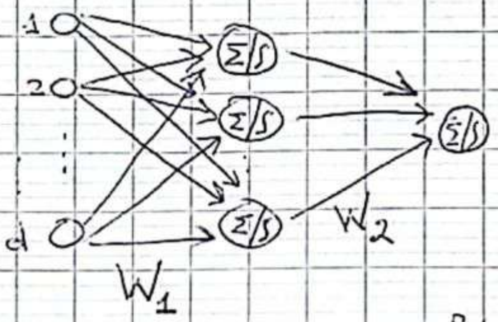
$$f(x) = \sigma_2(W_2\, \sigma_1(W_1 x))$$

որտեղ $W_1$-ը $d_1 \times d$ մատրից է

$\sigma_1$-ը $\mathbb{R}^{d_1} \to \mathbb{R}^{d_1}$ ոչ գծային արտապատկերում

$W_2 \in \mathbb{R}^{d_1}$

$\sigma_2 : \mathbb{R} \to \mathbb{R}$




$x \longmapsto f(x).$

<u>Դիտողություն՝</u> $Y$-ը կարող է նաև լինել բազմաչափ: Այդ դեպքում $Y \subset \mathbb{R}^{d_2}$, $d_2 > 1$, և $W_2$-ը կլինի $d_2 \times d_1$ չափանի մատրից: Օրինակ՝ պատկերների դեպքում, որտեղ լավ ընտրություն է $Y \in \mathbb{R}^K$ և

$$Y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{եթե անգղ.}$$

$$Y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{եթե ձի.}$$

և այլն. ∎

* Մի քանի թաքնված շերտով ՆՑ՝

$$f(x) = \underset{\substack{\downarrow \\ \text{ոչ գծային}}}{\sigma_L}\Big(\underbrace{W_L}_{\text{գծային}} \cdots \underset{\substack{\downarrow \\ \text{ոչ գծային}}}{\sigma_2}\Big(W_2 \sigma_1 \big(\underbrace{W_1 x}_{\text{գծային}}\big)\Big)\Big)$$

---

$\sigma_l$-ը անվանում ենք ակտիվացման ֆունկցիա: Ամենատարածված օրինակներն են՝

$$\sigma(x_1,\dots,x_d) = \begin{bmatrix} \mathbb{1}(x_1 \geq 0) \\ \vdots \\ \mathbb{1}(x_d \geq 0) \end{bmatrix}$$ 0-1 ակտիվացում

$$\sigma(x_1,\dots,x_d) = \begin{bmatrix} \max(x_1,0) \\ \vdots \\ \max(x_d,0) \end{bmatrix}$$ ReLU-ակտիվացում

$$\sigma(x_1,\dots,x_d) = \begin{bmatrix} \dfrac{\exp(x_1)}{1+\exp(x_1)} \\ \vdots \\ \dfrac{\exp(x_d)}{1+\exp(x_d)} \end{bmatrix}$$ լոգիստիկ կամ սիգմոիդ

$$\sigma(x_1,\dots,x_d) = \begin{bmatrix} \exp(x_1) / \sum_{j=1}^{d} \exp(x_j) \\ \vdots \\ \exp(x_d) / \sum_{j=1}^{d} \exp(x_j) \end{bmatrix}$$

Նեյրոնային ցանցի կառուցվածք ասելով հասկանում ենք հետևյալ հաջորդականությունը՝

$$(L, d_0, d_1, \dots, d_L)$$

| $L$ | $d_0$ | $d_1, \dots$ | $d_L$ |
|---|---|---|---|
| շերտերի քանակ կամ խորություն | մուտքի չափողականություն | շերտի լայնություն | ելքի չափողականություն |

Նեյրոնային ցանցի պարամետրեր ասելով հասկա-

Նշանակենք $\theta = (W_1, \dots, W_L)$ հավաքածու-
նությունը:

Այսուհետև, $f(x)$-ի փոխարեն հաճախ կգրենք $f(x;\theta)$:

Մարզել նեյրոնային ցանց, նշանակում է գտնել $\theta$-ի այնպիսի արժեք՝ $\hat{\theta}$, որ $\hat{f} = f(\cdot, \hat{\theta})$-ն լինի լավ կանխատեսիչ:

3) <u>Խորը ուսուցման մաթեմատիկական մարտահրավերները</u>

ՆՑ-ն մարզելու ամենատարածված մոտեցումն է՝ լուծել հետևյալ օպտիմիզացիայի խնդիրը

$$(*) \quad \hat{\theta} \in \arg\min_{\theta} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left(Y_i - f(X_i, \theta)\right)^2}_{:= \hat{L}_n(\theta) \text{ մարզման սխալանք}}$$

<u>Մարտահրավեր 1:</u>

$\hat{L}_n(\theta)$-ն ոչ ուռուցիկ ֆունկցիա է, որոշված ՆՑ-ի բարձր չափողականություն ունեցող բազմության վրա: Ինչպե՞ս լուծել $(*)$ խնդիրը:

## Մարտահրավեր II

Ինչպե՞ս ընտրել ՆՑ-ի կառուցվածքը, որ (*)-ի լուծում $\hat{\theta}$-ը ունենա ոչ միայն մարզման փոքր սխալանք, այլ նաև փոքր սխալանք չդիտարկված օրինակների վրա:

Էքսպերիմենտները ցույց են տալիս, որ I և II մարտահրավերները հաղթահարվում են կիրառելով գրադիենտի վայրէջք՝ շատ խորը կառուցվածք ունեցող ցանցերի վրա: Այդ ցանցերը գերպարամետրացված են (այսինքն $\theta$-ի չափողականությունը շատ ավելի մեծ է քան տվյալների քանակը): Հետևաբար, գոյություն ունեն $\theta$-ի բազմաթիվ (անվերջ քանակով) արժեքներ, որոնց համար $\hat{L}_n(\theta)=0$:

## Մարտահրավեր III

Ինչպե՞ս բացատրել այն երևույթը, որ գրադիենտի վայրէջքը խորը և գերպարամետրացված ցանցերի վրա ընտրում է այնպիսի $\hat{\theta}$, որի սխալանքը փոքր է: