

## Գրադիենտային վայրեջքի մեթոդ

Գրադիենտային վայրեջքի մեթոդը իտերացիոն ալգորիթմ է, որի յուրաքանչյուր քայլում կշիռների  $w$  վեկտորը փոփոխվում է նպատակային ֆունկցիայի առավելագույն նվազման, այսինքն՝ հակադրադիենտի ուղղությամբ՝

$$w = w - \eta \nabla Q,$$

որտեղ  $\eta$ -ն ուսուցման արագություն կոչվող պարամետրն է : Ուսուցման արագության պարամետրի (learning rate) ընտրությունն իրականացվում է փորձնական ճանապարհով, կախված է տվյալների բնույթից, որոնց համար ընտրվում է այդ պարամետրի արժեքը: Եթե օպտիմալացվող ֆունկցիան բավականին հարթ է, ապա ուսուցման արագությունը կարելի է վերցնել բավականին մեծ, որի հետևանքով ուսուցանող ալգորիթմի աշխատանքի ժամանակը կլինի փոքր: Բայց եթե նպատակային ֆունկցիան ունի մեծ կորություն, ապա ուսուցման մեծ արագության դեպքում հնարավոր է այնպիսի իրավիճակ, երբ կշիռների վեկտորը կարող է «թռնել» օպտիմալ արժեքի վրայով: Այդ դեպքերի համար պետք է վերցնել ուսուցման փոքր արագություն, մինիմումին զուգամիտման արագությունը նույնպես կլինի փոքր, բայց փոքր է նաև այն բաց թողնելու հավանականությունը:

Ուսուցանող ալգորիթմի իրականացման երկու մոտեցում կա.

1. փաթեթային (batch) — ալգորիթմի յուրաքանչյուր իտերացիայում ուսուցանող ընտրանին դիտարկվում է ամբողջովին, որից հետո միայն հաշվվում է  $w$  վեկտորի նոր արժեքը,
2. պատահական/ստոխաստիկ (stochastic) — ալգորիթմի յուրաքանչյուր իտերացիայում պատահականորեն ընտրվում է ուսուցանող ընտրանուց միայն մեկ օբյեկտ:

Առաջին մոտեցումն ավելի բարդ է, բայց այն ավելի արագ է զուգամիտում շնորհիվ այն բանի, որ դիտարկվում է ամբողջ ընտրանին:

Երկրորդ մոտեցումը ենթադրում է ավելի պակաս հաշվումներ, բայց պահանջում է ուսուցման համար հաջորդ տարրի ընտրության ալգորիթմի մշակում:

Փաթեթային մեթոդն իրականացնելը շատ պարզ է, անհրաժեշտ է ուսուցանող ընտրանու բոլոր օրինակների հակագրադիենտի միջինացված արժեքների միջոցով իրականացնել կշիռների թարմացում այնքան ժամանակ, քանի դեռ  $Q$  ֆունկցիոնալի գնահատականը չի կայունացել, այսինքն՝ մինչև նրա արժեքը կսկսի փոփոխվել շատ դանդաղ կամ ընդհանրապես չի փոփոխվի, կամ քանի դեռ փոփոխվում են կշիռները: Կշիռների թարմացման գործընթացը կարելի է ներկայացնել հետևյալ առնչություններով՝

$$W_0 = W_0 - \eta \frac{1}{n} \sum_{i=1}^n (wx_i - y_i)$$

$$W_j = W_j - \eta \frac{1}{n} \sum_{i=1}^n (wx_i - y_i) x_{i,j}$$

Հաշվողական ռեսուրսների օպտիմալ օգտագործման տեսանկյունից առավել հետաքրքիր է պատահական գրադիենտի ալգորիթմը, որի համար որպես մուտք հանդես են գալիս  $XL$  ուսուցանող ընտրանին,  $\eta$  -ուսուցման արագությունը,  $Q$  -ի հարթեցման  $\lambda$  պարամետրը: Ալգորիթմի ելքը կշիռների  $w$  վեկտորն է:

Պատահական գրադիենտի մեթոդի ալգորիթմը հետևյալն է՝

1. Տալ կշիռների սկզբնական վեկտորը  $w$ :
2. Որոշել  $Q$  ֆունկցիոնալի ընթացիկ գնահատականը

$$Q(w) = \sum_{i=1}^n (a(x_i) - y(x_i))^2$$

3. Կրկնել, քանի դեռ  $Q$  -ի արժեքը չի կայունացել և (կամ)  $w$  կշիռները չեն դադարել փոփոխվել:

3.1.  $XL$ -ից ընտրել  $X_j$ :

3.2. Հաշվել ալգորիթմի ելքային արժեքը՝  $a(w, X_j)$ -ն և սխալանքը

$$\varepsilon_i = (a(x_i) - y(x_i))^2$$

3.3. Կատարել գրադիենտային վայրէջքի ուղղությամբ քայլ(որոշել կշիռների նոր արժեքները)

$$w = w - 2\eta(wx_i - y_i)x_i$$

4. Գնահատել  $Q$  ֆունկցիոնալի արժեքը՝

$$Q = (1 - \lambda)Q + \lambda \varepsilon_i$$

Կա կշիռների սկզբնական վեկտորի տրման մի քանի միջոց.

1) կշիռների վեկտորը ներկայացնել որպես զրոյական վեկտոր,

2) կշիռների վեկտորը տալ պատահական թվերի միջոցով  $[-\frac{1}{k}; \frac{1}{k}]$  միջակայքից, որտեղ  $k$  -ն հայտանիշների տարածության չափայնությունն է:

Պատահական գրադիենտի մեթոդի առավելությունը փաթեթայինի նկատմամբ նաև այն է, որ այդ ալգորիթմը հնարավորություն է տալիս իրականացնել այսպես կոչված օնլայն ուսուցում, և եթե գծային ռեգրեսիայի դեպքում այդ առանձնահատկությունը էական չէ, ապա դասակարգման խնդիրների լուծման ժամանակ, մանավանդ տարբեր բանական վեբ ծառայությունների աշխատանքի պայմաններում, այն հնարավորություն է տալիս ալգորիթմին վերապատրաստվել իրական ժամանակում: Պատահական գրադիենտի մեթոդը, ի տարբերություն փաթեթային մոտեցման, հնարավորություն է տալիս ուսուցման համար օգտագործել գերմեծ ծավալի ընտրանիներ: Բացի այդ, այս մեթոդում կարելի է փոփոխել հենց ուսուցման ալգորիթմը, եթե ընտրանին շատ մեծ է, ապա օբյեկտը դիտարկել ու հետո այն կարելի է հեռացնել ընտրանուց և դրանով օպտիմալացնել համակարգչի ռեսուսների օգտագործումը: Եթե ընտրանին շատ փոքր է, կարելի է օբյեկտները ընտրել վերադարձով:

Պատահական գրադիենտի մեթոդի թերություններն են.

- Հնարավոր է ալգորիթմի դանդաղ զուգամիտում:

- Եթե  $Q(w)$  ֆունկցիոնալն ունի մեծ թվով էքստրեմալ կետեր, ապա մեծ է հավանականությունը, որ ալգորիթմը կժխա լոկալ մինիմումի վրա: Դրա դեմ պայքարելու համար օգտագործում են  $w$  վեկտորի պատահական ձևափոխության մեթոդը, որի արդյունքում վեկտորի մեկ կամ մի քանի կոորդինատ մի քիչ փոփոխվում է, դրանով փոքրացնելով լոկալ մինիմումի տիրույթում հայտնվելու հավանականությունը: