L1 և L2 Ռեգուլյարիզացիաներ

Հայկ Կարապետյան

Ռեգուլյարիզացիան գերուսուցումից խուսափելու մեթոդը: Ծանոթանանք, թե ինչպես է դա հնարավոր: Ունենք տան գնի գուշակման խնդիր: Սահմանենք նեյրոնային ցանց բաղկացած մեկ շերտից և k հատ նեյրոնից: Պատկերացնենք ֆունկցիաների բազմություն (A), որոնք ստացվել $f(x) = wx + b$ ֆունկցիայի մեջ տարբեր w և b արժեքներ տեղադրելիս:

$$A = \{wx + b, w \in R^k,\ b \in R\}$$

Գերուսուցված ֆունկցիան և լավագույն ֆունկցիան, որը կարողանում է լուծել տան գնի գուշակման խնդիրը 100% ճշգրտությամբ, նույնպես գտնվում են այդ բազմության մեջ: Մոդելը ուսուցանելիս մենք ձգտում ենք գտնել այդ լավագույն ֆունկցիան, բայց կարող ենք հասնել գերուսուցված ֆունկցիան: Այդ պատճառով եկեք փոքրացնենք մեր բազմությունը: Եկեք սահմանափակենք w-ի փոփոխման տիրույթը: Օրինակ՝ w-ն ձգտեցնենք զրոյի: Դա հնարավոր է կազմակերպել օգտագործելով կորստի ֆունկցիան: Այսինքն մեզ պետք է կատարել այնպիսի ձևափոխություն կորստի ֆունկցիայի մեջ, որ w-ն ձգտի 0-ի:

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

$$L_{L1} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 + \frac{\lambda}{k}\sum_{i=1}^{k}|w_i|$$

$$L_{L2} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 + \frac{\lambda}{k}\sum_{i=1}^{k}w_i^2$$

Այստեղ λ-ն հիպերպարամետր է, որը ցույց է տալիս, թե մեզ համար ինչքան կարևոր է կշիռների փոքր լինելը, k-ն շերտում նեյրոնների քանակը: λ-ի արժեքը փոփոխվում է (0; 1) միջակայքում, քանի որ $f(x) \approx y$ ավելի կարևոր է, քան կշիռների փոքր լինելը: Կշիռները փոքր լինելու միջոցով, մոդելի գերուսուցումից խուսափելու մեթոդը կոչվում է ռեգուլյարիզացիա (regualrization): Երկու տեսակի ռեգուլյարիզացիաներն էլ ապահովում են կշիռների փոքր լինելը, բայց նրանց մեջ կա տարբերություն: Դիտարկենք այդ տարբերությունը կշիռների թարմացման ժամանակ:

$$L1\ Regularization$$

$$w_1 = w_1 - \alpha \times (\nabla L + \frac{\partial \frac{\lambda}{k}\sum_{i=1}^{k}|w_i|}{\partial w_1})$$

$$w_1 = w_1 - \alpha \times \frac{w_1}{|w_1|} - \alpha\nabla L$$

$$w_1 = w_1 - \alpha \times signum(w_1) - \alpha\nabla L$$

Ամեն անգամ w_1-ին գումարում ենք կամ հանում ենք α

Եվ դրա շնորհիվ հնարավոր է w_1-ի արժեքը հասցնել 0-ի

$$L2\ Regularization$$

$$w_1 = w_1 - \alpha \times (\nabla L + \frac{\partial \frac{\lambda}{k}\sum_{i=1}^{k}w_i^2}{\partial w_1})$$

$$w_1 = w_1 - \alpha \times 2w_1 - \alpha\nabla L$$

$$w_1 = w_1(1 - 2\alpha) - \alpha\nabla L$$

w_1-ից ամեն անգամ դեն ենք նետում իր $(1 - 2\alpha)$ մասը և այն 0 չի դառնա

L1 ռեգուլյարիզացիան հայտնի է նաև Lasso regression անվանումով, իսկ L2 ռեգուլյարիզացիան Ridge regression անվանումով: Ինչպես նշեցինք, երկուսն էլ գերուսուցումից խուսափելու մեթոդներ են, բայց դա չի նշանակում, որ դրանք օգտագործելիս մոդելը չի կարող գերուսուցվել:

L1 ռեգուլարիզացիան նաև օգտագործվում է անպետք հատկանիշներ (features) գտնելու համար: Օրինակ՝ տան գների գուշակման խնդրումկատարել ենք ռեգրեսիա՝ L1 ռեգուլարիզացիայով: Ուսուցումից հետո մոդելի կշիռները տպելիս հնարավոր է, որ մի քանիսը շատ մոտ լինեն 0-ի: Դա նշանակում է, որ տվյալ հատկանիշի արժեքը չի ազդում տան գնի վրա (տան տիրոջ անուն):