

Matemātika: Gradientu pārspīlējums

Problēmas apraksts

Savam pēdējam uzdevumam Ivans lūdza Alisi palīdzēt viņam ar "Zaļās linijas" (Green Line) ML uzdevuma īstenošanu. Konkrēti, viņš lūdza izpētīt gradienta lejupeja (Gradient Descent) algoritma matemātiskos pamatus. Palīdziet Alisei pabeigt viņas uzdevumu!

Uzdevums

Mēs mēģinām apmācīt sarežģītu modeli (1D lineārā regresija) uz sarežģītas datu kopas (1 1D punkts, kas atrodas pie $x = 5$), un mums ir nepieciešama Jūsu palīdzība...

Būtībā, šādai mazai dimensionalitātei problēma tiek samazināta līdz punkta meklēšanai ar lineāro regresijas modeli. Lai definētu, cik tuvu mēs esam, mēs izmantojam tā saukto zaudējumu funkciju (loss function). Šai problēmai tā ir definēta kā attāluma kvadrāts starp pašreizējo punktu un punktu $x = 5$:

$$L(x) = (x - 5)^2$$

Varat pārbaudīt, ka zaudējumu funkcijas vērtība ir minimāla un vienāda ar nulli pie $x = 5$.

Modelis sākas no patvaļīga punkta x_0 un mēģina sasniegt punktu $x = 5$. Katrā solī tas pārbauda, vai iešana pa labi palielina zaudējumu funkcijas vērtību. To dara, aprēķinot zaudējumu funkcijas L *slipumu* vai *atvasinājumu* attiecībā pret tās *argumentu* x , ko apzīmē ar $\frac{dL}{dx}$. Intuitīvi būtu jāpārvietojas uz pozīciju ar mazāku L vērtību. Kad iešana pa labi palielina zaudējumu funkciju (kas nozīmē, ka atvasinājums ir pozitīvs), ir jēga iet pa kreisi, un otrādi. Lai aprēķinātu, cik lielam solim jābūt, mēs ņemam atvasināto un reizinām to ar konstanti α . α izvēle parasti ir Jūsu ziņā (un, spoileru brīdinājums, Jūs izpētīsiet, kā izvēlēties šo konstanti). Pagaidām pieņemsim, ka α ir pozitīvs.

Formālāk, pāreja no soļa t uz $t + 1$ tiek dota ar formulu:

$$x_{t+1} = x_t - \alpha \cdot \frac{dL}{dx_t}$$

Atvasinājums tiek dots kā

$$\frac{dL}{dx_t} = 2 \cdot (x_t - 5)$$

Pārbaudiet, ka pozitīvam atvasinājumam solis tiek veikts pa kreisi un otrādi.

Ja Jūs esat vairāk pazīstami ar mašīnmācīšanos, Jūs, iespējams, atpazīsiet procedūru, kas pazīstama kā gradienta lejupeja (gradient descent), un α kā mācīšanās ātrumu (learning rate).

Jūsu uzdevums ir šāds:

{1} Izpildiet gradienta lejupeju, sākot no $x_0 = 0$, ar $\alpha = 0.3$. Varat apstāties, ja zaudējumu funkcijas L vērtība solī x_t ir ne vairāk kā 1. Katram solim norādiet: $x_t, L(x_t), \frac{dL}{dx_t}, x_{t+1}$.

{2} Izpētiet, kā gradienta lejupeja norisinās dažādām α vērtībām. Pamanīsiet, ka modelis tuvojas 5 noteiktām α vērtībām, bet citām — nē. Aprakstiet, kurām α vērtībām modelis konverģē (vai nekonverģē) uz 5. Identificējiet α vērtību diapazonu, kuram gradienta lejupeja konverģē uz 5. Noteikti analizējiet atrastā diapazona robežpunktus. Skaidrojumam jābūt pēc iespējas pilnīgākam un jāietver visu iespējamo gadījumu analīze.

Norāde: meklējot atbildi, Jums var noderēt ģeometrisku rindu īpašības un robežas.

Formāts

Uzdevumam {1} iesniedziet tabulu ar kolonnām: t , x_t , $L(x_t)$, $\frac{dL}{dx_t}$, x_{t+1} .

Uzdevumam {2} sniedziet matemātisku skaidrojumu. Neprasām formālu matemātisku pierādījumu, bet pilniem punktiem nepieciešams derīgs matemātisks skaidrojums par α intervāla atvasināšanu, ne tikai novērojumi.

Iekļaujiet būtiskās detaļas, bet izvairieties no liekas plašsaistes (ja nesniedzat formālu pierādījumu, mēģiniet ietilpināt risinājumu līdz pusei A4 lapas). Tas palīdzēs ātrāk un precīzāk novērtēt Jūsu risinājumu. Jūs varat sniegt savu risinājumu latviešu, krievu vai angļu valodā.

Vērtēšana

- 25% par pareizu {1}
- 75% par pareizu {2}

Daļēji punkti ir iegūstami abās daļās par daļēji pareizām atbildēm.

Papildu paskaidrojums

Ja Jūs vēlaties uzzināt vairāk par gradienta lejupeju un tās izmantošanu mašīnmācīšanās, mēs iesakām:

- īsu video no Visually Explained: <https://www.youtube.com/watch?v=qg4PchTECck>
- garāku skaidrojumu no 3b1b (daļa no viņa neironu tīklu sērijas): <https://www.youtube.com/watch?v=IHZwWFHWa-w>