VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS KATEDRA

MLP mokymas naudojant sujungtinių gradientų metodą

Referatas

Atliko: Rytis Karpuška (parašas)

Vilnius – 2017

TURINYS

1. SUJUNGTINIŲ GRADIENTŲ METODAS 2
   1.1. Sujungtiniai vektoriai 2
   1.2. F(x) optimizavimas 2
   1.3. F(x) iteracinis optimizavimas 3
      1.3.1. Žemiausio taško juostoje radimas 4

IŠVADOS 6

1. Sujungtinių gradientų metodas

Daugiasluoksnio perceptrono mokymui labai svarbus yra nuostolių funkcijos $\epsilon(x)$ minimizavimas, kuris padeda rasti įeities svorius neuronams. Sujungtinių gradientų metodas yra algoritmas skirtas optimizuoti antro laipsnio lygčių sistemoms (1) kurios gali aproksimuoti arba tiksliai išreikšti neuronų tinklo nuostolių funkciją.

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c \tag{1}$$

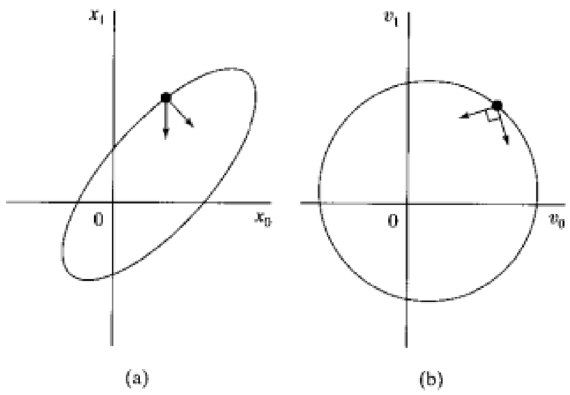
Čia x yra stulpelinis vektorius, kurio ilgį žymėsime w, t.y. $w = ||x||$, o A kvadratinė w x w, simetriška, teigiamai apibrėžta ($z^T Az > 0$ visiems nenuliniams z vektoriams) matrica.

1.1. Sujungtiniai vektoriai

Sujungtinių gradientų metodas remiasi sujungtiniais vektoriais. Šie vektoriai $s_1, ..., s_w$ nedaro vienas kitam įtakos matricos A atžvilgiu t.y.:

$$s_i^T As_j = 0 \text{ visiems } i \neq j \tag{2}$$

Verta pastebėti, kad jeigu A yra vienetinė matrica, tai sujungtiniai vektoriai sutampa su statmenais.



1 pav. sujungtiniai vektoriai.

Panagrinėkime pavyzdį pateiktą 1 pav. A dalyje pateikta elipsė atitinkanti dvimatį (1) variantą. Taip pat pavaizduoti du sujungtiniai vektoriai. Jeigu matricą A transformuosime taip, kad elipsė taptų apskritimu - sujungtiniai veiktoriai taps statmeni.

1.2. F(x) optimizavimas

Sujungtinių krypčių metodas turimiems $s_1, ..., s_w$ apibrėžiamas kaip

$$x_{n+1} = x_n + \eta_n s_n, \ n = 1, ..., w \tag{3}$$

kur x_0 yra bet koks pradinis vektorius, o η yra apibrėžiama kaip

$$f(x_n + \eta_n s_n) = \min_{\eta} f(x_n + \eta s_n) \tag{4}$$

Neformaliai kalbant, sujungtinių krypčių metodas pradeda bet kuriame taške x_0 ir atlieka $f(x)$ reikšmių "pjūvį" s_0 kryptimi. Tuomet tame "pjūvyje" randa mažiausią $f(x)$ reikšmę ir pasirenką tą tašką, kaip naują pradžios tašką. Toliau tas pats kartojama su likusiais s vektoriais. Verta pastebėti, kad sujungtinių gradientų metodas reikauja ne daugiau kaip w sujungtinių vektorių minimumo radimui.

1.3. F(x) iteracinis optimizavimas

Jau 1.2 poskyryje buvo netiesiogiai užsiminta, apie tai, kad šis metodas taikomas iteraciškai. Sujungtinių vektorių radimui apsibrėšime stačiausio nuolydžio (atvirkštinės išvestinės) krypties funkciją.

$$r_n = b - Ax_n \tag{5}$$

Vykdant iteracijas, vektorius s apibrėžiamas kaip stačiausio nuolydžio ir prieš tai buvusio sujungtinio vektoriaus tiesinė kombinacija.

$$s_n = r_n + \beta_n s_{n-1} \tag{6}$$

pradinis sujungtinis vektorius prilyginamas stačiausio nuolydžio krypčiai:

$$s_0 = r_0 \tag{7}$$

koeficiento β_n radimui galima naudoti tiesiogiai išvestą formulę priklausiančią nuo A, tačiau Polak–Ribière, Fletcher–Reeves, Hestenes-Stiefel, Dai–Yuan pasiūlė efektyvesnių variantų. Čia panagrinėsime Polak–Ribière, Fletcher–Reeves, pasiūlytus variantus.

Polak–Ribière formulė. Polak–Ribière pasiūlė tokį variantą β_n radimui:

$$\beta_n = \frac{r_n^T(r_n - r_{n-1})}{r_{n-1}^T r_{n-1}} \tag{8}$$

Fletcher–Reeves formulė. Fletcher–Reeves pasiūlė tokį variantą β_n radimui:

$$\beta_n = \frac{r_n^T r_n}{r_{n-1}^T r_{n-1}} \tag{9}$$

Abi šios formulės nenaudoja matricos A dėl ko jų efektyvumas palyginus su įprastais metodais yra geresnis. Taip pat jos abi yra ekvivaliančios kvadratinei $f(x)$. Tačiau jeigu jos yra naudojamos atvejams kuomet nuostolių funkcija nėra kvadratinė, bet aproksimuojama (pvz.: atmetant aukštesnio laipsnio dalis) - generuojami s vektoriai nėra tiksliai sujungtiniai ir po keletos iteracijų gali užstrigti. Dažnai renkamasi Polak–Ribière formulė su apribojimu:

$$\beta_n = max[\beta_{pr}, 0] \tag{10}$$

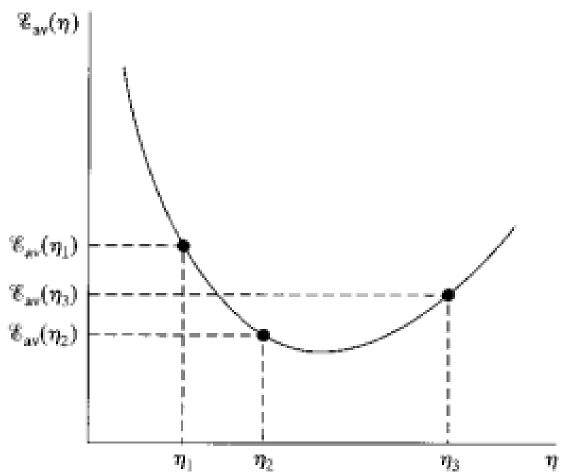
kur β_{pr} yra Polak–Ribière β iteracijoje n.

1.3.1. Žemiausio taško juostoje radimas

Turint s_n vektorių reikia rasti minimalią $\epsilon(x)$ funkcijos reikšmę jo kryptimi. Pateikiamas numerinis būdas rasti $\epsilon(x)$ minimalią reikšmę. Iš pradžių pasirenkame tris taškus taip, kad

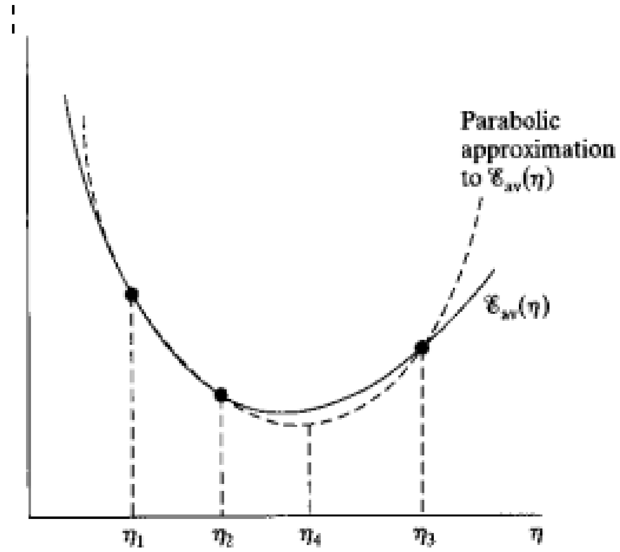
$$\epsilon(\eta_1) \geq \epsilon(\eta_3) \geq \epsilon(\eta_2), \text{ kur } \eta_1 < \eta_2 < \eta_3 \tag{11}$$

Tuomet ant šių taškų užbrėžiame kvadratinė kreivę taip, kad ji eitų per visus tris taškus (išsprendžiame kvadratinę lygtį) (žr.: 2 pav).



2 pav. Kvadratine kreivė einanti per tris pasirinktus taškus prieš patikslinimą.

Tuomet randame šios kvadratinės funkcijos minimumą, ir minimumo tašką priskiriame η_4. η_1, η_2 ir η_4 naudojame kaip naujus tris taškus minimizavimui. Toliau kartojame procesą, kol rasta minimumo reikšmė tenkins pasirinktus baigties kriterijus (žr.: 3 pav).



3 pav. Kvadratine kreivė einanti per tris pasirinktus taškus po patikslinimo.

Išvados

Šiame referate buvo aprašytas sujungtinių gradientų metodas dažnai naudojamas įvairiose optimizavimo bei mašinų mokymo uždaviniuose.