Legmeredekebb ereszkedés (steepest descent)

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

2023. november 14.

Függvényextrémum keresés: többváltozós eset

Többváltozós $f(\mathbf{x})$ függvény esetén az egydimenziós esetben hasznos módszer (pl. minimum bekeretezése) általában nem használható

A módszerek most is két csoportra oszthatóak:

- csak függvénykiértékelések szükségesek, pl Nelder–Mead-módszer
- a függvény $\nabla f(\mathbf{x})$ gradiensét is fel tudjuk használni, pl **legmeredekebb ereszkedés** módszer

A legmeredekebb ereszkedés

Induljunk ki egy tetszőleges $\mathbf{P}_0$ pontból

- határozzuk meg a gradiensvektort: $\nabla f(\mathbf{P}_0)$
- $-\nabla f(\mathbf{P}_0)$ által kijelölt **irány** mentén csökken a leggyorsabban a függvény
- a következő lépés a fenti irány által kijelölt **egyenes mentén** történő minimalizáció
- ehhez a lépéshez használhatjuk az egydimenziós esetre kifejlesztett módszereket
- így érünk el a $\mathbf{P}_1$ minimum pontba, ez általában egy lokális minimum
- a lépéssor ismétlése $-\nabla f(\mathbf{P}_i)$ irányba
- ez a legmeredekebb ereszkedés (steepest descent) módszer

A legmeredekebb ereszkedés

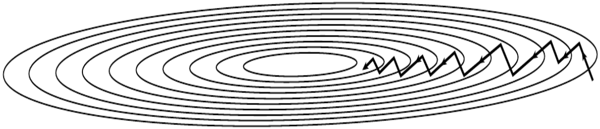

Figure: Legmeredekebb ereszkedés iteráció. ©Numerical Recipies

Probléma

- az újabb minimum pontban a gradiens mindig merőleges a előző irányra
- ez abból következik, hogy (lokális) minimumba értünk (ha nem ez lenne a helyzet, akkor az azt jelentené, hogy az előző irány menti derivált nem nulla a minimumban)
- "szűk völgyekben" nagyon lassan halad az iteráció
- erre lehet megoldás az ún. **konjugált gradiens módszer**

Legmeredekebb ereszkedés regressziós feladatokban[1]

Gépi tanulással kapcsolatos feladatokban gyakran a legmeredekebb ereszkedés egy egyszerűbb változatát használják

- adott költségfüggvény $J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$, ahol **a** paramétervektor
- keressük $\min_{\mathbf{a}}[J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})]$

[1] Andrew Ng, Coursera alapján

Legmeredekebb ereszkedés regressziós feladatokban[1]

Gépi tanulással kapcsolatos feladatokban gyakran a legmeredekebb ereszkedés egy egyszerűbb változatát használják

- adott költségfüggvény $J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$, ahol $\mathbf{a}$ paramétervektor
- keressük $\min_{\mathbf{a}}[J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})]$

Stratégia (gradient descent)

- válasszunk egy tetszőleges kezdőértéket az $\mathbf{a} = (a_0, a_1 \ldots a_k)^T$ vektornak
- `repeat until convergence` $\{$
  $a_j := a_j - \alpha \frac{\partial}{\partial a_j} J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$
  $\}$
- $\alpha > 0$: *learning parameter*, nekünk kell megadni

[1] Andrew Ng, Coursera alapján

Legmeredekebb ereszkedés regressziós feladatokban

Stratégia:

$$
\begin{aligned}
&\mathtt{a} = \mathtt{a}^{(init)} \\
&\ \mathtt{repeat\ until\ convergence}\ \{ \\
&a_j := a_j - \alpha \frac{\partial}{\partial a_j} J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)}) \\
&\}
\end{aligned}
$$

- a legmeredekebb ereszkedés algoritmussal ellentétben nincs irány menti minimalizáció a legközelebbi lokális minimumig
- viszont bevezetjük α paramétert
- ha α túl nagy, akkor elvétjük a minimumot, ha túl kicsi, lassú a konvergencia
- mit jelent a konvergencia figyelése? később tárgyaljuk

Legmeredekebb ereszkedés regressziós feladatokban

Kérdés: függ-e az interáció $\mathbf{a}^{(final)}$ eredménye attól, hogy milyen $\mathbf{a}^{(init)}$ kezdőértéket választunk?

- általában függhet, ha a $J(\mathbf{a}; \mathbf{x}^{(i)}, y^{(i)})$ költségfüggvény olyan, hogy több lokális minimuma van
- az iteráció beragadhat egy közeli lokális minimumba