Rješavanje optimizacijskih problema algoritmima evolucijskog računanja u Javi
Numeričke optimizacije. Uspješnost generiranja početnih rješenja.

dr.sc. Marko Čupić

Fakultet elektrotehnike i računarstva
Sveučilište u Zagrebu
Akademska godina 2013./2014.

10. listopada 2013.

Layout

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

$$\frac{d^2f}{dx^2} = 6$$

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

$$\frac{d^2 f}{dx^2} = 6$$

- Kako izgleda graf ove funkcije?

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

$$\frac{d^2f}{dx^2} = 6$$

- Kako izgleda graf ove funkcije?
- Gdje je minimum?

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

$$\frac{d^2f}{dx^2} = 6$$

- Kako izgleda graf ove funkcije?
- Gdje je minimum? Za $f' = 0$.

Prvi primjer

Pretpostavimo da je zadana funkcija $f(x) = 3x^2 - 6x - 45$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad skalarom $x \in \mathcal{R}$.

- Funkcija ima prvu derivaciju definiranu u svim točkama.

$$\frac{df}{dx} = 6x - 6$$

- Funkcija ima drugu derivaciju definiranu u svim točkama.

$$\frac{d^2f}{dx^2} = 6$$

- Kako izgleda graf ove funkcije?
- Gdje je minimum? Za $f' = 0$. $6 \cdot x_{min} - 6 = 0 \rightarrow x_{min} = 1$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u minimumu?

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u minimumu?

$$\left.\frac{df}{dx}\right|_{x=1} = 6x - 6|_{x=1} = 6 - 6 = 0$$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u $x > 1$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u $x > 1$

$$\left.\frac{df}{dx}\right|_{x>1} = 6x - 6|_{x>1} > 0$$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u $x > 1$

$$\left.\frac{df}{dx}\right|_{x>1} = 6x - 6|_{x>1} > 0$$

- Koja je vrijednost u $x < 1$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Koja je vrijednost u $x > 1$

$$\left.\frac{df}{dx}\right|_{x>1} = 6x - 6|_{x>1} > 0$$

- Koja je vrijednost u $x < 1$

$$\left.\frac{df}{dx}\right|_{x<1} = 6x - 6|_{x<1} < 0$$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, povećanjem x-a vrijednost funkcije će se također povećati.

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, povećanjem x-a vrijednost funkcije će se također povećati.
- Ako je $f'(x) < 0$, smanjenjem x-a vrijednost funkcije će se također povećati.

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, x ćemo smanjiti kako bi vrijednost funkcije pala: $x \leftarrow x - \delta$, $\delta > 0$.

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, x ćemo smanjiti kako bi vrijednost funkcije pala: $x \leftarrow x - \delta$, $\delta > 0$.
- Ako je $f'(x) < 0$, x ćemo povećati kako bi vrijednost funkcije pala: $x \leftarrow x + \delta$, $\delta > 0$.

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, x ćemo smanjiti kako bi vrijednost funkcije pala: $x \leftarrow x - \delta$, $\delta > 0$.
- Ako je $f'(x) < 0$, x ćemo povećati kako bi vrijednost funkcije pala: $x \leftarrow x + \delta$, $\delta > 0$.
- Ako je $f'(x) \neq 0$, možemo pisati $x \leftarrow x - \lambda \cdot f'(x)$, $\lambda > 0$.

Prvi primjer

Predznak prve derivacije u nekoj točki x definira što treba napraviti toj točki da bi vrijednost funkcije porasla.

$$f' = \frac{df}{dx} = 6x - 6$$

Za traženje minimuma funkcije tada ćemo okrenuti pravilo: potrebno je kretati se u smjeru negativne derivacije!

- Ako je $f'(x) = 0$, vrijednost x je lokalni optimum (ili točka infleksije, ili ...).
- Ako je $f'(x) > 0$, x ćemo smanjiti kako bi vrijednost funkcije pala: $x \leftarrow x - \delta$, $\delta > 0$.
- Ako je $f'(x) < 0$, x ćemo povećati kako bi vrijednost funkcije pala: $x \leftarrow x + \delta$, $\delta > 0$.
- Ako je $f'(x) \neq 0$, možemo pisati $x \leftarrow x - \lambda \cdot f'(x)$, $\lambda > 0$.

Za koliki iznos δ to smijemo promijeniti x?

Layout

Drugi primjer

Pretpostavimo da je zadana funkcija $f(\vec{x}) = (x_1 - 4)^2 + (x_2 + 8)^2 + (x_3 + 5)^2$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad n-dimenzijskim vektorom $\vec{x} \in \mathcal{R}^3$.

- Više ne možemo računati *derivaciju* jer je to funkcija od više varijabli.

Drugi primjer

Pretpostavimo da je zadana funkcija $f(\vec{x}) = (x_1 - 4)^2 + (x_2 + 8)^2 + (x_3 + 5)^2$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad n-dimenzijskim vektorom $\vec{x} \in \mathcal{R}^3$.

- Više ne možemo računati *derivaciju* jer je to funkcija od više varijabli.
- Možemo računati gradijent – pisat ćemo ga kao jednostupčani vektor čiji je element u i-tom retku parcijalna derivacija zadane funkcije po i-toj varijabli odnosno i-toj komponenti vektora $\vec{x}$.

Drugi primjer

Pretpostavimo da je zadana funkcija $f(\vec{x}) = (x_1 - 4)^2 + (x_2 + 8)^2 + (x_3 + 5)^2$ čiji tražimo minimum. To je skalarna funkcija koja je definirana nad n-dimenzijskim vektorom $\vec{x} \in \mathcal{R}^3$.

- Više ne možemo računati *derivaciju* jer je to funkcija od više varijabli.
- Možemo računati gradijent – pisat ćemo ga kao jednostupčani vektor čiji je element u i-tom retku parcijalna derivacija zadane funkcije po i-toj varijabli odnosno i-toj komponenti vektora $\vec{x}$.

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

Drugi primjer

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

Drugi primjer

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

- Funkcija poprima ekstremnu vrijednost (ili sedlo, ili ...) u točkama u kojima je $\nabla f(\vec{x}) = \vec{0}$.

Drugi primjer

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

- Funkcija poprima ekstremnu vrijednost (ili sedlo, ili ...) u točkama u kojima je $\nabla f(\vec{x}) = \vec{0}$.
- Rezultira sustavom jednadžbi koji nije uvijek moguće riješiti egzaktno $\rightarrow$ vodi na iterativne postupke koji malo po malo mijenjaju vrijednost trenutnog rješenja $\vec{x}$.

Drugi primjer

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

- Funkcija poprima ekstremnu vrijednost (ili sedlo, ili ...) u točkama u kojima je $\nabla f(\vec{x}) = \vec{0}$.
- Rezultira sustavom jednadžbi koji nije uvijek moguće riješiti egzaktno $\rightarrow$ vodi na iterativne postupke koji malo po malo mijenjaju vrijednost trenutnog rješenja $\vec{x}$.
- Interpretacija gradijenta je međutim slična kao i interpretacija derivacije: promjena vrijednosti $\vec{x}$ za skalirani iznos $\lambda \cdot \nabla f(\vec{x})$, $\lambda > 0$, u okolici od $\vec{x}$ vodi ka povećanju vrijednosti funkcije.

Drugi primjer

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 - 8 \\ 2x_2 + 16 \\ 2x_3 + 10 \end{bmatrix}$$

- Funkcija poprima ekstremnu vrijednost (ili sedlo, ili ...) u točkama u kojima je $\nabla f(\vec{x}) = \vec{0}$.
- Rezultira sustavom jednadžbi koji nije uvijek moguće riješiti egzaktno $\rightarrow$ vodi na iterativne postupke koji malo po malo mijenjaju vrijednost trenutnog rješenja $\vec{x}$.
- Interpretacija gradijenta je međutim slična kao i interpretacija derivacije: promjena vrijednosti $\vec{x}$ za skalirani iznos $\lambda \cdot \nabla f(\vec{x})$, $\lambda > 0$, u okolici od $\vec{x}$ vodi ka povećanju vrijednosti funkcije.
- Za minimizaciju stoga biramo korekciju proporcionalnu iznosu $-\lambda \cdot \nabla f(\vec{x})$. Za koliki λ se smijemo pomaknuti?

Layout

Koliko je prikladno?

- U slučaju funkcije od jedne varijable došli smo do ažuriranja:

$$x \leftarrow x - \lambda \cdot f'(x)$$

Koliko je prikladno?

- U slučaju funkcije od jedne varijable došli smo do ažuriranja:

$$x \leftarrow x - \lambda \cdot f'(x)$$

- U slučaju funkcije od više varijabli došli smo do ažuriranja:

$$\vec{x} \leftarrow \vec{x} - \lambda \cdot \nabla f(\vec{x})$$

Koliko je prikladno?

- U slučaju funkcije od jedne varijable došli smo do ažuriranja:

$$x \leftarrow x - \lambda \cdot f'(x)$$

- U slučaju funkcije od više varijabli došli smo do ažuriranja:

$$\vec{x} \leftarrow \vec{x} - \lambda \cdot \nabla f(\vec{x})$$

- U oba slučaja ostalo je neodgovoreno za koliko se smijemo pomaknuti?

Koliko je prikladno?

- U slučaju funkcije od jedne varijable došli smo do ažuriranja:

$$x \leftarrow x - \lambda \cdot f'(x)$$

- U slučaju funkcije od više varijabli došli smo do ažuriranja:

$$\vec{x} \leftarrow \vec{x} - \lambda \cdot \nabla f(\vec{x})$$

- U oba slučaja ostalo je neodgovoreno za koliko se smijemo pomaknuti?
    - Ako je $\lambda > 0$ premali, da smo uzeli veći, mogli smo funkciju više minimizirati u trenutnom koraku.

Koliko je prikladno?

- U slučaju funkcije od jedne varijable došli smo do ažuriranja:

$$x \leftarrow x - \lambda \cdot f'(x)$$

- U slučaju funkcije od više varijabli došli smo do ažuriranja:

$$\vec{x} \leftarrow \vec{x} - \lambda \cdot \nabla f(\vec{x})$$

- U oba slučaja ostalo je neodgovoreno za koliko se smijemo pomaknuti?
    - Ako je $\lambda > 0$ premali, da smo uzeli veći, mogli smo funkciju više minimizirati u trenutnom koraku.
    - Ako je $\lambda > 0$ preveliki, izlazimo iz *okolice* od $\vec{x}$ pa tako daleko funkcija se može ponašati drugačije (npr. rasti).

Algoritmi pretraživanja u zadanom smjeru?

U oba prethodna slučaja zadana je točka (x ili $\vec{x}$) te smjer u kojem se treba pomaknuti ($-\lambda \cdot f'(x)$ ili $-\lambda \cdot \nabla f(\vec{x})$) – potrebno je pronaći prikladnu vrijednost za λ.

Algoritmi pretraživanja u zadanom smjeru?

U oba prethodna slučaja zadana je točka (x ili $\vec{x}$) te smjer u kojem se treba pomaknuti ($-\lambda \cdot f'(x)$ ili $-\lambda \cdot \nabla f(\vec{x})$) – potrebno je pronaći prikladnu vrijednost za λ.

- Postoji niz algoritama koji se mogu koristiti za rješavanje ovog problema.

Algoritmi pretraživanja u zadanom smjeru?

U oba prethodna slučaja zadana je točka (x ili $\vec{x}$) te smjer u kojem se treba pomaknuti ($-\lambda \cdot f'(x)$ ili $-\lambda \cdot \nabla f(\vec{x})$) – potrebno je pronaći prikladnu vrijednost za λ.

- Postoji niz algoritama koji se mogu koristiti za rješavanje ovog problema.
- Rijetko kada imamo egzaktno rješenje – najčešće se radi o iterativnim postupcima.

Algoritmi pretraživanja u zadanom smjeru?

U oba prethodna slučaja zadana je točka (x ili $\vec{x}$) te smjer u kojem se treba pomaknuti ($-\lambda \cdot f'(x)$ ili $-\lambda \cdot \nabla f(\vec{x})$) – potrebno je pronaći prikladnu vrijednost za λ.

- Postoji niz algoritama koji se mogu koristiti za rješavanje ovog problema.
- Rijetko kada imamo egzaktno rješenje – najčešće se radi o iterativnim postupcima.
- Kako je višedimenzijski slučaj općenitiji ($\vec{x}$), njega ćemo razmotriti.

Algoritmi pretraživanja u zadanom smjeru?

Zadana je funkcija $f(\vec{x})$ čiji se traži minimum, pri čemu je $\vec{x} \in \mathcal{R}^n$. Neka je $\vec{x}^{(0)}$ odabrano početno rješenje. Pretpostavimo također da je $\vec{d}^{(0)} \in \mathcal{R}^n$ vektor koji pokazuje u smjeru u kojem je potrebno modificirati trenutno rješenje kako bi vrijednost funkcije pala, gledano iz trenutne točke $\vec{x}^{(0)}$.
Uvodimo parametar λ te definiramo novu funkciju:

$$\theta(\lambda) = f(\vec{x}^{(0)} + \lambda \cdot \vec{d}^{(0)}).$$

Funkcija θ ovisi samo o *lambda*; za nju su $\vec{x}^{(0)}$ i $\vec{d}^{(0)}$ konstante. Sada je zadatak pronaći vrijednost λ^* koja minimizira funkciju $\theta(\lambda)$ i potom kao novo rješenje uzeti vrijednost:

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \lambda^* \cdot \vec{d}^{(0)}.$$

Algoritmi pretraživanja u zadanom smjeru?

Neka je $f(\vec{x}) = (x_1 - 4)^2 + (x_2 + 8)^2 + (x_3 + 5)^2$, $\vec{x}^{(0)} = [6, -10, -9]^T$ te $\vec{d} = [-1, 2, 1]^T$. Lagano se možemo uvjeriti da će za male vrijednosti λ funkcija padati te će potom u jednom trenutku početi rasti:

λ	$\vec{x}^{(0)} + \lambda \cdot \vec{d}^{(0)}$	$\theta(\lambda)$
0	$[6, -10, -9]^T$	24
1	$[5, -8, -8]^T$	10
2	$[4, -6, -7]^T$	8
3	$[3, -4, -6]^T$	18
4	$[2, -2, -5]^T$	40

Algoritmi pretraživanja u zadanom smjeru?

Pogledajmo

$$\frac{d\theta(\lambda)}{d\lambda} = \nabla f(\vec{x})^T \cdot \vec{d}.$$

Algoritmi pretraživanja u zadanom smjeru?

Pogledajmo

$$\frac{d\theta(\lambda)}{d\lambda} = \nabla f(\vec{x})^T \cdot \vec{d}.$$

- Vrijedi

$$\left.\frac{d\theta(\lambda)}{d\lambda}\right|_{\lambda=0} < 0.$$

Zašto to vrijedi? Definirat ćemo stoga da je $\lambda_{lower} = 0$.

Algoritmi pretraživanja u zadanom smjeru?

Pogledajmo

$$\frac{d\theta(\lambda)}{d\lambda} = \nabla f(\vec{x})^T \cdot \vec{d}.$$

- Vrijedi

  $$\left.\frac{d\theta(\lambda)}{d\lambda}\right|_{\lambda=0} < 0.$$

  Zašto to vrijedi? Definirat ćemo stoga da je $\lambda_{lower} = 0$.
- Trebamo gornju ogradu za λ, oznaka λ_{upper}, uz koju će funkcija f ponovno početi rasti, odnosno za koju će $\frac{d\theta(\lambda)}{d\lambda_{upper}} > 0$.

Algoritmi pretraživanja u zadanom smjeru?

Pogledajmo

$$\frac{d\theta(\lambda)}{d\lambda} = \nabla f(\vec{x})^T \cdot \vec{d}.$$

- Vrijedi

  $$\left.\frac{d\theta(\lambda)}{d\lambda}\right|_{\lambda=0} < 0.$$

  Zašto to vrijedi? Definirat ćemo stoga da je $\lambda_{lower} = 0$.
- Trebamo gornju ogradu za λ, oznaka λ_{upper}, uz koju će funkcija f ponovno početi rasti, odnosno za koju će $\frac{d\theta(\lambda)}{d\lambda_{upper}} > 0$.
  - Ako takvu nemamo, možemo isprobavati $\lambda_{upper} = 1$, $\lambda_{upper} = 2$, $\lambda_{upper} = 4$, $\lambda_{upper} = 8$, ... dok ne pronađemo potreban λ_{upper}.

Algoritmi pretraživanja u zadanom smjeru?

Jednom kada imamo λ_{lower} i λ_{upper}, sigurno znamo da se λ koji minimizira funkciju nalazi između te dvije ograde. Možemo koristiti, primjerice, binarno raspolavljanje.

Metoda bisekcije

Korak 0. Postavi $k = 0$. Postavi $\lambda_l = \lambda_{lower}$ te $\lambda_u = \lambda_{upper}$.

Korak k. Postavi $\lambda = \frac{\lambda_l + \lambda_u}{2}$ i izračunaj $\left.\frac{d\theta(\lambda)}{d\lambda}\right|_\lambda$.

Ako je $\left.\frac{d\theta(\lambda)}{d\lambda}\right|_\lambda > 0$, redefiniraj $\lambda_u = \lambda$, $k = k + 1$.

Ako je $\left.\frac{d\theta(\lambda)}{d\lambda}\right|_\lambda < 0$, redefiniraj $\lambda_l = \lambda$, $k = k + 1$.

Ako je $\left.\frac{d\theta(\lambda)}{d\lambda}\right|_\lambda \approx 0$, stani jer smo došli dovoljno blizu.

Postoji još niz drugih sličnih metoda.

Algoritam pretraživanja u zadanom smjeru

Jednom kada znamo kako pronaći prikladan korak, možemo definirati porodicu algoritama pretraživanja u zadanom smjeru (engl. *line-search algorithms*).

Pretraživanje u zadanom smjeru.

Ponavljaj za $k = 1, 2, \ldots$

Ako je $\vec{x}^{(k)}$ optimum, prekini s izvođenjem i vrati $\vec{x}^{(k)}$.

Inače

Utvrdi $\vec{d}^{(k)}$ – smjer pretrage

Pronađi $\lambda^* > 0$ – korak

Definiraj $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda^* \cdot \vec{d}^{(k)}$ – novo rješenje

Algoritam gradijentnog spusta

Metoda gradi linearni model funkcije oko trenutnog rješenja. Kao vektor $\vec{d}^{(k)}$ bira upravo minus gradijent funkcije koju je potrebno minimizirati.

$$\vec{d}^{(k)} = -\nabla f(\vec{x})$$

Algoritam gradijentnog spusta

Metoda gradi linearni model funkcije oko trenutnog rješenja. Kao vektor $\vec{d}^{(k)}$ bira upravo minus gradijent funkcije koju je potrebno minimizirati.

$$\vec{d}^{(k)} = -\nabla f(\vec{x})$$

- U ne baš rijetkim slučajevima, izračun gradijenta je računski skupa operacija!

Algoritam gradijentnog spusta

Metoda gradi linearni model funkcije oko trenutnog rješenja. Kao vektor $\vec{d}^{(k)}$ bira upravo minus gradijent funkcije koju je potrebno minimizirati.

$$\vec{d}^{(k)} = -\nabla f(\vec{x})$$

- U ne baš rijetkim slučajevima, izračun gradijenta je računski skupa operacija!
- Procjena optimalne vrijednosti λ^* može tražiti puno izračuna gradijenta funkcije; stoga se ponekad pomak radi i uz manje "dobru" vrijednost za λ.

Algoritam gradijentnog spusta

Metoda gradi linearni model funkcije oko trenutnog rješenja. Kao vektor $\vec{d}^{(k)}$ bira upravo minus gradijent funkcije koju je potrebno minimizirati.

$$\vec{d}^{(k)} = -\nabla f(\vec{x})$$

- U ne baš rijetkim slučajevima, izračun gradijenta je računski skupa operacija!
- Procjena optimalne vrijednosti λ^* može tražiti puno izračuna gradijenta funkcije; stoga se ponekad pomak radi i uz manje "dobru" vrijednost za λ.
- U najjednostavnijim varijantama za vrijednost λ se koristi mala pozitivna konstanta (npr. $\lambda = 0.1$ ili čak $\lambda = 0.01$) i ne provodi se pretraga za λ^*.

Algoritam gradijentnog spusta

Metoda gradi linearni model funkcije oko trenutnog rješenja. Kao vektor $\vec{d}^{(k)}$ bira upravo minus gradijent funkcije koju je potrebno minimizirati.

$$\vec{d}^{(k)} = -\nabla f(\vec{x})$$

- U ne baš rijetkim slučajevima, izračun gradijenta je računski skupa operacija!
- Procjena optimalne vrijednosti λ^* može tražiti puno izračuna gradijenta funkcije; stoga se ponekad pomak radi i uz manje "dobru" vrijednost za λ.
- U najjednostavnijim varijantama za vrijednost λ se koristi mala pozitivna konstanta (npr. $\lambda = 0.1$ ili čak $\lambda = 0.01$) i ne provodi se pretraga za λ^*.
- Primjer je vrlo poznati algoritam za učenje neuronskih mreža: *Backpropagation*.

Newtonova metoda

Metoda gradi kvadratni model $g(\vec{x})$ funkcije $f(\vec{x})$ oko trenutnog rješenja.

$$f(\vec{x}+\vec{\tau}) = f(\vec{x}) + \nabla f(\vec{x})^T \cdot \vec{\tau} + \frac{1}{2!}\vec{\tau}^T \nabla^2 f(\vec{x})\vec{\tau} + \zeta \qquad (1)$$

pri čemu su s ζ označene pogreške višeg reda. Za potrebe izgradnje Newtonovog algoritma u okolici točke $\vec{x}$ funkcija $f(\vec{x})$ modelira se kvadratnom funkcijom $g(\vec{x})$ koja je definirana na sljedeći način:

$$g(\vec{x}+\vec{\tau}) = f(\vec{x}) + \nabla f(\vec{x})^T \cdot \vec{\tau} + \frac{1}{2!}\vec{\tau}^T \nabla^2 f(\vec{x})\vec{\tau}. \qquad (2)$$

$g(\vec{x}+\vec{\tau})$ pri tome promatramo kao funkciju od $\vec{\tau}$ i pitamo se koji $\vec{\tau}$ treba odabrati da bismo dobili minimum?

Newtonova metoda

$H(f) = \nabla^2 f(\vec{x})$ je Hesseova matrica odnosno matrica drugih parcijalnih derivacija. Ova matrica računa se na sljedeći način:

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{3}$$

Newtonova metoda

Derivacijom

$$g(\vec{x}+\vec{\tau}) = f(\vec{x}) + \nabla f(\vec{x})^T \cdot \vec{\tau} + \frac{1}{2!}\vec{\tau}^T \nabla^2 f(\vec{x})\vec{\tau}$$

po $\vec{\tau}$ slijedi:

$$\nabla g = \nabla f(\vec{x}) + H(\vec{x}) \cdot \vec{\tau}. \tag{4}$$

Uvjet za optimum je da je gradijent jednak 0:

$$\nabla f(\vec{x}) + H(\vec{x}) \cdot \vec{\tau} = \vec{0} \tag{5}$$

iz čega slijedi:

$$H(\vec{x}) \cdot \vec{\tau} = -\nabla f(\vec{x}) \tag{6}$$

odnosno

$$\vec{\tau} = -H(\vec{x})^{-1} \cdot \nabla f(\vec{x}) \tag{7}$$

Newtonova metoda

- Izraz $H(\vec{x})^{-1}$ predstavlja matrični inverz Hesseove matrice i upravo potreba za računom inverza čini ovaj postupak računski vrlo zahtjevnim.
- Dobiveni vektor $\vec{\tau}$ dalje se koristi kao vektor u čijem smjeru se radi pretraga prethodno opisanim algoritmom.

$$\vec{d}^{(k)} = \vec{\tau} = -H(\vec{x})^{-1} \cdot \nabla f(\vec{x})$$

- Treba odmah uočiti: ako je funkcija f kvadratna funkcija, tada joj njezin model g savršeno odgovara i postupak minimizacije će završiti u jednom koraku.

Razlog za neuspjeh?

Postupak stvaranja početnog rješenja može ne-uspjeti; najčešći razlog je kršenje tvrdih ograničenja. Jedan od prokušanih načina koji može pomoći:

Savjeti

- Pratite neuspjehe.
- Temeljem povijesti neuspjeha procijenjujte težinu (problematičnost; zahtjevnost) sastavnih dijelova rješenja.
- Dajte zahtjevnijim dijelovima rješenja priliku da prvi budu razriješeni.
- Izbor možete učiniti determinističkim ili vjerojatnosnim (teža komponenta, veća šansa da se ranije razriješi).

Pogledajte primjer i diskusiju u knjizi u poglavlju 5.3.