34 (A4B33OPT)

1 Metoda nejmenších čtverců

Metoda nejmenších čtverců je matematicko-statistická metoda používaná zejména při zpracování nepřesných dat (typicky experimentálních empirických dat získaných například měřením). Metoda je v základní podobě určená pro řešení nekompatibilních soustav lineárních rovnic (v obecnější podobě hovoříme o nelineární metodě nejmenších čtverců), díky čemuž je fakticky ekvivalentní tzv. lineární regresi.

Řešme nehomogenn soustavu linearnch rovnic Ax = b

A je matice o rozměrech m×n a soustava má řešení právě tehdy, když b∈rngA, jinak je soustava přeurčená (typicky, když m>n, víc rovnic než neznámých)

V našem případě řešíme přeurčenou soustavu, tedy hledáme takové x, že vzdálenost mezi body Ax a b je co nejmenší, tedy:

$min\{||Ax - b|| \mid x \in R^n\}$

Místo normy klidně můžeme minimalizovat její čtverec:

$||Ax - b||^2$

1.1 Použití na regresi:

Regrese je modelování zavislosti proměnné y∈R na proměnné t∈T regresn funkcí y = f(t, x), která je známá, až na parametry $x \in R^n$ Je dán soubor dvojic (t_i, y_i), i = 1, ... ,m, kde měření $y_i \in R$ jsou zatížena chybou. Úkolem je najít parametry x, aby y_i f(t$_i$, x). Podle metody nejmenších čtverců tedy řešme.

$$\min_{x \in R^n} \sum_{i=1}^{m} (f(t_i, x) - y_i)^2$$

2 Analytické podmínky na lokální extrémy

2.1 Volné extrémy

V tomto případě hledáme lokální extrémy funkce.

Máme dva důležité typy bodů:

1. stacionární bod - bod, kde je funkce diferencovatelná a všechny parciální derivace jsou nulové.

2. kritický bod - bod, který je buď stacionární nebo v něm není funkce diferencovatelná.

2.1.1 Podmínka prvního řádu

Všechny kritické body jsou podezřelé z volného lokálního extrému.[1]

[1] Když počítáme příklad nejdříve uděláme všechny parciální derivace, následně je položíme rovny nule a vyřešíme soustavu rovnic => kritické body

2.1.2 Podmínky druhého řádu

- f má v bodě x ostré lokální minimum [maximum] na X právě tehdy, když Hessova matice druhých derivací f"(x) je pozitivně[2] [negativně[3]] definitní.
- Je-li f"(x) indefinitní[4], nemá f v x lokální minimum ani lokální maximum na X.
- Je-li f"(x) pozitivně [negativně] semidefinitní, nevíme o tomto bodě jestli je minimem, maximem nebo ani jedno z toho.

2.2 Vázané extrémy

V tomto případě hledáme lokální extrémy funkce za určité podmínky dané nejčastěji jinou funkcí nebo funkcemi.

2.2.1 postup

Extrémy hledáme za pomoci lagrangeových multiplikátorů λ, řešíme rovnici:
$f'(x) + \lambda g'(x) = 0^T$[5]

3 Numerické metody pro optimalizaci bez omezení

U všech dále zmíněných případů se jedná o iterační numerické metody pro nalezení volného lokálního minima diferencovatelných funkcí na množině R^n

3.1 Gradientní metoda

Metoda volí směr sestupu jako záporný gradient funkce f v bode x_k[6].
Tedy $x_{k+1} = x_k - \alpha_k (A^T A)^{-1} f'(x_k)^T$
Rychlost konvergence bývá často pomalá, kvůli cik-cak chování.

3.2 Newtonova metoda

Newtonova metoda je iteracn algoritmus na resen soustav nelinearnch rovnic. Lze ho pouzt i na minimalizaci funkce tak, ze hledame nulovy gradient.

---

[2] V příkladech hledáme vlastní čísla matice, když jsou všechna > 0, pak je poz. def.

[3] V příkladech hledáme vlastní čísla matice, když jsou všechna < 0, pak je neg. def.

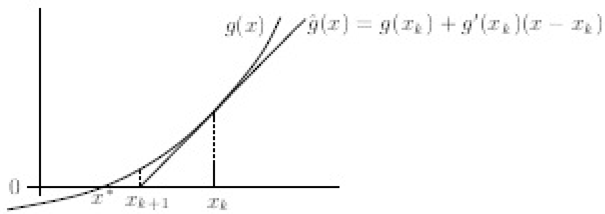
[4] V příkladech hledáme vlastní čísla matice, když existuje vlastní číslo, které < 0 a zároveň existuje vlastní číslo, které >0, pak je matice indefinitní

[5] Tedy v praxi si napíšeme zadání a podmínky si převedeme do tvaru, kdy je na jedné straně rovnice nula a roznásobíme je λ_1 *až* λ_n, tento výraz se nazývá Lagrangeova funkce, ze které spočítáme první derivace a vyřešíme soustavu rovnic, z níchž dostaneme body podezřelé z extrémů. (zjištění druhu extrému je podle skript i wernera složité a v OPT vůbec nebylo)

[6] tj. vždy jdeme nejstrmějším směrem dolů

3.2.1 Použití na soustavy nelineárních rovnic

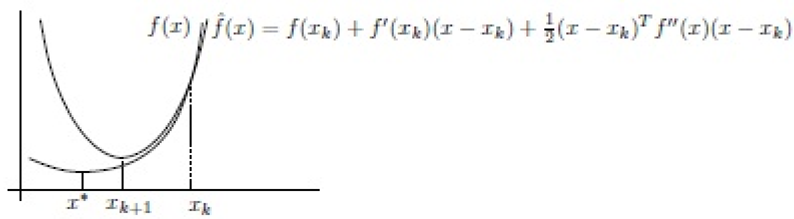
Zobrazení g aproximujeme v okolí bodu x_k taylorovým polynomem prvního řádu $g(x) = g(x_k) + g'(x_k)(x - x_k)$



Při řešení soustavy rovnic najdeme další krok, jako $x_{k+1} = x_k - g'(x_k)^{-1}g(x_k)$

3.2.2 Použití na hledání lokálního minima

V tomto případě aproximujeme funkci f taylorovým polynomem druhého řádu.



Tedy newtonovu metodu lze použít pro hledání lokálního extremu dvakrát diferencovatelné funkce, když položíme $g(x) = f'(x)^T$, z toho dostaneme, že iterace:

$x_{k+1} = x_k - f''(x_k)^{-1}f'(x_k)^T$

3.3 Gaussova-Newtonova metoda

Snažíme se najít přibližné řešení ve smyslu nejmenších čtverců, což vede na minimalizaci funkce:

$f(x) = ||g(x)||^2 = g(x)^T g(x)$

Další krok hledáme následovně:

$x_{k+1} = x_k - (g'(x_k)^T g'(x_k))^{-1} g'(x_k)^T g(x_k)$, což se dá v případě, že $g'(x_k)$ má plnou hodnost napsat, jako $x_{k+1} = x_k - g'(x_k)^+ g(x_k)$, kde $g'(x_k)^+$je pseudoinverze.

Výhody: Vyhneme se počítání druhých derivací (hesiánů)

Nevýhody: Metoda má horší konvergenční chování než Newtonova metoda

3.4 Levenberg-Marquardtova metoda

Levenbergova-Marquardtova metoda je široce používané vylepšení Gaussovy-Newtonovy metody, které matici $g'(x_k)^T g'(x_k)$ nahrazuje maticí $g'(x_k)^T g'(x_k) +$

$\mu_k I$, kde μ_k>0

Tedy další krok hledáme následovně:

$x_{k+1} = x_k - (g'(x_k)^T g'(x_k) + \mu_k I)^{-1} g'(x_k)^T g(x_k)$

Zajímavosti:

- Pro malé μ_k se Levenbergova-Marquardtova metoda blíží Gauss-Newtonově metodě.
- Pro velké μ_k se Levenbergova-Marquardtova metoda blíží Gradientní metodě s délkou kroku ${\mu_k}^{-1}$.

Parametr μ_k měníme po každé iteraci.

- Pokud iterace snížila účelovou funkci, pak iteraci přijmeme a μ_k zmenšíme
- Pokud iterace nesnížila účelovou funkci iteraci zamítneme a μ_k zvětšíme

Zmenšování azvětšování μ_k děláme násobením a dělením konstantou.