Prednáška 3

Moderné gradientové optimalizačné metódy pre funkcie n-premenných

Študujme **reálnu funkciu** n-premenných

$$f\colon R^n \to R$$

Našou úlohou bude nájsť také $x_{opt} \in R^n$, pre ktoré má funkcia f minimum

$$\mathbf{x}_{opt} = \arg\min f(\mathbf{x})$$

Túto úlohu budeme riešiť pomocou:

(1) Newtonovou metódou
(2) Metódou združených gradientov (conjugated gradients)
(3) Metódou premennej metriky (variable metrics)

Posledné dve metódy sa v literatúre označujú ako "**moderné** (gradientové) optimalizačné metódy".

1. Newtonova metóda

Jednoduché zovšeobecnenie Newtonovej optimalizačnej metódy pre funkcie 1-premennej na prípad funkcií n-premenných.

Budeme približne riešiť rovnicu

$$\operatorname{grad} f(\boldsymbol{x}) = \boldsymbol{0}$$

Nech $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{\delta}$, kde $\boldsymbol{\delta}=(\delta_1,\delta_2,\ldots,\delta_n)$, potom

$$\operatorname{grad} f(\boldsymbol{x}_o + \boldsymbol{\delta}) = \boldsymbol{0}$$

Ľavá strana bude upravená pomocou Taylorovho rozvoja

$$\operatorname{grad} f(\boldsymbol{x}_o) + H(\boldsymbol{x}_o)\boldsymbol{\delta} + \ldots = \boldsymbol{0}$$

Ak budeme uvažovať len prvé dva členy rozvoja a budeme predpokladať, že Hessián $H(\boldsymbol{x}_o)$ je nesingulárna matica, potom

$$\boxed{\boldsymbol{\delta} \approx -H^{-1}(\boldsymbol{x}_o)\operatorname{grad} f(\boldsymbol{x}_o)}$$

Poznámka: Výchylka $\boldsymbol{\delta}$ je formálne určená dvoma ekvivalentnými spôsobmi: buď ako riešenie systému lineárnych rovníc

$$H(\boldsymbol{x}_o)\boldsymbol{\delta} \approx -\mathrm{grad}\, f(\boldsymbol{x}_o)$$

alebo pomocou inverznej matice

$$\boldsymbol{\delta} \approx -H^{-1}(\boldsymbol{x}_o)\,\mathrm{grad}\, f(\boldsymbol{x}_o)$$

Pretože, numerický výpočet inverznej matice ja približne dvakrát náročnejší ako priame riešenie systému lineárnych rovníc, doporučuje sa vektor $\boldsymbol{\delta}$ určiť riešením systému lineárnych rovníc. Explicitný výraz pre $\boldsymbol{\delta}$ pomocou inverznej matice budeme používať z dôvodov formálnej jednoduchosti zápisu.

Riešenie rovnice $\operatorname{grad} f(\boldsymbol{x}) = \boldsymbol{0}$ má potom tento približný tvar

$$\boxed{\boldsymbol{x} \approx \boldsymbol{x}_o - H^{-1}(\boldsymbol{x}_o)\operatorname{grad} f(\boldsymbol{x}_o)}$$

Toto riešenie slúži ako podklad pre konštrukciu rekurentnej formule, ktorá je základom Newtonovej optimalizačnej metódy

$$\boxed{\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H^{-1}(\boldsymbol{x}_k)\operatorname{grad} f(\boldsymbol{x}_k)}$$

pre k=0,1,2,... a $\boldsymbol{x}_\circ$ je zadaná počiatočné riešenie. Rekurentná aplikácia tejto formule je ukončená keď začne platiť $|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k| < \varepsilon$, kde ε>0 je zadaná presnosť požadovaného riešenia.

Algoritmus Newtonovej metódy:

```
read(x₀,kmax,ε);
k:=0; norm:=∞; x:=x₀;
while (k<kmax) and (norm>ε) do
begin k:=k+1;
      solve SLE H(x)δ=-grad f(x);
      x':=x+δ;
      norm:=|x-x'|;
      x:=x';
end;
write(x,f(x));
```

Príkaz solve SLE $H(\mathbf{x})\delta=-\text{grad } f(\mathbf{x})$; znamená, že použitím nejakej metódy (napr. Gaussovej eliminačnej metódy) systém lineárnych rovníc (SLE) $H(\mathbf{x})\delta=-\text{grad } f(\mathbf{x})$ je vyriešený vzhľadom k vektoru δ.

Poznámky:

(1) Newtonova metóda nájde riešenie rovnice $\operatorname{grad} f(\boldsymbol{x}) = \boldsymbol{0}$, t.j. nájde stacionárne stavy funkcie $f(\boldsymbol{x})$. Tieto stacionárne stavy sú klasifikované pomocou Hessiánu $H(\boldsymbol{x})$:

> Ak $\boldsymbol{x}$ je stacionárny bod a Hessián $H(\boldsymbol{x})$ je pozitívne negatívny, potom v bode $\boldsymbol{x}$ má funkcia $f(\boldsymbol{x})$ minimum.

(2) Pri Newtonovej metóde nemáme nijakú záruku či konverguje k minimu alebo k maximu (konverguje k stacionárnemu bodu).

(3) Funkcia $f(\boldsymbol{x})$ musí byť dvakrát diferen-covateľná na celom R^n, počítame gradient a Hessián funkcie.

(4) Výpočet Hessiánu môže byť časovo a pamäťovo veľmi náročný, menovite pre funkcie mnohých (niekoľko sto) premenných.

1a. Linearizovaná Newtonova metóda

Pre určitý typ minimalizovanej funkcie výpočet Hessiánu môže byť podstatne zjednodušený.

$$f(\boldsymbol{x}) = \sum_{k=1}^{p} g_k^2(\boldsymbol{x})$$

Pre túto funkciu platí vlastnosť

$$f\left(\boldsymbol{x}^*\right) = 0 \quad \Leftrightarrow \quad \forall k \in [1, p]: g_k\left(x^*\right) = 0$$

Nutná a postačujúca podmienka pre existenciu takého $\boldsymbol{x}^*$, $f(\boldsymbol{x}^*)=0$, aby pre každé $k \in [1, p]$ platilo $g_k(\boldsymbol{x}^*)=0$.

Funkcia f(x) sa vyskytuje v **regresnej analýze** a pri transformácii konštrukcie **riešenia systému nelineárnych rovníc na minimali-začný problém.**

Regresná analýza ako minimalizačný problém

Majme regresnú tabuľku (tréningovú množinu)

$$A_{train} = \left\{ \boldsymbol{x}_i \;/\; y_i ; i = 1,2,\ldots,p \right\}$$

Našou úlohou je adaptovať parametre $\boldsymbol{\omega}$ modelovej funkcie $F(\boldsymbol{x},\boldsymbol{\omega})$ tak, aby čo najlepšie reprodukovala tréningovú množinu

$$E(\boldsymbol{\omega}) = \sum_{i=1}^{p} \left(F\left(\boldsymbol{x}_i, \boldsymbol{\omega}\right) - y_i \right)^2$$

Adaptačný proces modelovej funkcie spočíva v tom, že sa minimalizuje účelová funkcia $E(\boldsymbol{\omega})$

$$\boldsymbol{\omega}_{opt} = \arg\min_{\boldsymbol{\omega}} E(\boldsymbol{\omega})$$

Transformácia konštrukcie riešenie systému nelineárnych rovníc na minimalizačný problém.

Študujme systém nelineárnych rovníc

$$g_1(x_1, x_2, \ldots, x_n) = 0$$
$$g_2(x_1, x_2, \ldots, x_n) = 0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$g_p(x_1, x_2, \ldots, x_n) = 0$$

Našou úlohou je nájsť také $\boldsymbol{x}^*$ pre ktoré vyššie uvedené rovnice budú splnené. Definujeme si funkciu

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{k=1}^{p} g_k^2(\boldsymbol{x})$$

Ak sa nám podarí nájsť také globálne minimum $\boldsymbol{f}(\boldsymbol{x}^*) = 0$, potom vektor $\boldsymbol{x}^*$ je riešením systému nelineárnych rovníc.

Výpočet Hessiánu pre funkciu

$$f(\boldsymbol{x}) = \sum_{k=1}^{p} g_k^2(\boldsymbol{x})$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = 2\sum_{k=1}^{p} g_k(\boldsymbol{x}) \frac{\partial g_k(\boldsymbol{x})}{\partial x_i}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = 2\sum_{k=1}^{p} \frac{\partial g_k(\boldsymbol{x})}{\partial x_i} \frac{\partial g_k(\boldsymbol{x})}{\partial x_j} + \cancel{2\sum_{k=1}^{p} g_k(\boldsymbol{x}) \frac{\partial^2 g_k(\boldsymbol{x})}{\partial x_i \partial x_j}}$$

Budeme predpokladať, že vektor $\boldsymbol{x}$ je blízko $\boldsymbol{x}^*$, potom $g_k(\boldsymbol{x}) \approx 0$, pre k=1,2,..., potom približný výraz pre elementy Hessiánu má tento tvar

$$\boxed{H_{ij}(\boldsymbol{x}) \approx \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \approx 2\sum_{k=1}^{p} \frac{\partial g_k(\boldsymbol{x})}{\partial x_i} \frac{\partial g_k(\boldsymbol{x})}{\partial x_j}}$$

V maticovom formalizme tento výraz má tvar

$$\boxed{H(\boldsymbol{x}) \approx 2\sum_{k=1}^{p} \left(\text{grad } g_k(\boldsymbol{x})\right)^T \left(\text{grad } g_k(\boldsymbol{x})\right)}$$

Newtonová metóda v ktorej

(1) účelová funkcia má tvar

$$f(\boldsymbol{x}) = \sum_{k=1}^{p} g_k^2(\boldsymbol{x})$$

(2) a Hessián sa počíta pomocou gradientu

sa nazýva **linearizovaná Newtonova metóda.** Obvykle znamená podstatné **urýchlenie** algoritmu.