Általános lineáris függvényillesztés

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

2022. február 07.

Általános lineáris függvényillesztés

Az $\{(x^{(i)}, y^{(i)})\}$ adathalmazra szeretnénk illeszteni egy modellt. Most is a

$$\chi^2(\mathbf{a}; x^{(i)}, y^{(i)}) = \sum_{i=1}^{N} \left[\frac{y^{(i)} - h(x^{(i)}; \mathbf{a})}{\sigma_i} \right]^2$$

költségfüggvényt minimalizáljuk.

$h(x; \mathbf{a})$ alakja:

$$h(x; \mathbf{a}) = \sum_{j=1}^{M} a_j f_j(x),$$

- $f_j(x)$ tetszőleges ún. bázisfüggvény
- $f_j(x)$ nem függ az a_j-től

Általános lineáris függvényillesztés

A minimumban

$$\frac{\partial \chi^2(\mathbf{a})}{\partial a_j} = 0$$

Behelyettesítve $h(x; \mathbf{a})$-t:

$$\frac{\partial \chi^2(\mathbf{a})}{\partial a_j} = 2 \cdot \sum_{i=1}^{N} \left[\frac{1}{\sigma_i^2} \cdot \left(\sum_{k=1}^{M} a_k f_k(x^{(i)}) - y^{(i)} \right) \cdot f_j(x^{(i)}) \right] = 0$$

Ez egy lineáris egyenletrendszer az a_k együtthatókra!

A tervmátrix[1]

Jelölések:

$$X_{ij} = \frac{f_j(x^{(i)})}{\sigma_i} \qquad\qquad b_i = \frac{y^{(i)}}{\sigma_i}$$

X_{ij} az úgynevezett *tervmátrix*:

- az M oszlopa a bázisfüggvényeknek felel meg
- az N sora a mérési pontoknak
- a mátrixelemek a j. bázisfüggvény $x^{(i)}$ helyeken vett értékei

[1]design matrix

A tervmátrix és a b_i vektor felépítése

$$X_{ij} = \begin{pmatrix} \frac{f_1(x^{(1)})}{\sigma_1} & \frac{f_2(x^{(1)})}{\sigma_1} & \dots & \frac{f_M(x^{(1)})}{\sigma_1} \\ \frac{f_1(x^{(2)})}{\sigma_2} & \frac{f_2(x^{(2)})}{\sigma_2} & \dots & \frac{f_M(x^{(2)})}{\sigma_2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{f_1(x^{(N)})}{\sigma_N} & \frac{f_2(x^{(N)})}{\sigma_N} & \dots & \frac{f_M(x^{(N)})}{\sigma_N} \end{pmatrix} \qquad b_i = \begin{pmatrix} \frac{y^{(1)}}{\sigma_1} \\ \frac{y^{(2)}}{\sigma_2} \\ \vdots \\ \vdots \\ \frac{y^{(N)}}{\sigma_N} \end{pmatrix}$$

A lineáris illesztés normálegyenletei

A parciális deriváltakra felírt egyenletek ezzel:

$$\sum_{i=1}^{N}\left[\left(\sum_{k=1}^{M} a_k X_{ik} - b_i\right)\cdot X_{ij}\right] = 0,$$

Indexes írásmóddal:

$$X_{ik}\,a_k X_{ij} = X_{ij}\,b_i$$

A lineáris illesztés normálegyenletei

A parciális deriváltakra felírt egyenletek ezzel:

$$\sum_{i=1}^{N}\left[\left(\sum_{k=1}^{M} a_k X_{ik} - b_i\right)\cdot X_{ij}\right] = 0,$$

Indexes írásmóddal:

$$X_{ik}\,a_k X_{ij} = X_{ij}\,b_i$$

Mátrixos írásmódban:

$$\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{a} = \mathbf{X}^{\mathsf{T}}\mathbf{b},$$

- $\mathbf{X}$ egy $N \times M$-es mátrix, tehát $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ egy $M \times M$-es mátrix
- $\mathbf{a}$ tartalmazza az ismeretleneket
- ez egy $M \times M$ lineáris egyenletrendszer: az illesztés normálegyenletei

Többváltozós polinomillesztés

További példa: $y^{(i)}$ több változótól függ: $\mathbf{x}^{(i)}$ K-dimenziós vektor

Ekkor definiáljuk a következő polinomot

$$h(\mathbf{x}; \mathbf{a}) = \sum_{k=1}^{K} \sum_{j=1}^{M} a_j^{(k)} [x_k]^{j-1},$$

- a probléma immár összesen $M \cdot K$ ismeretlent fog tartalmazni
- ebben még nincsenek vegyes tagok!
- a probléma ugyanúgy oldható meg, mint az előző

Normálegyenletek megoldása

Normálegyenleteket:

$$\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{a} = \mathbf{X}^\mathsf{T}\mathbf{b}$$

Hogyan oldjuk meg?

- lineáris egyenlet megoldó numerikus rutinok
- viszont az egyenletrendszer lehet rosszul kondicionált $\Rightarrow$ pontatlan megoldás
- ilyenkor lehet hasznos az ún szinguláris érték felbontás (singular value decomposition, SVD)

Normálegyenletek vs iteratív módszerek

A költségfüggvény minimalizálására **iteratív módszerek** is léteznek:

- pl legmeredekebb ereszkedés (gradient descent)
- és olyan módszerek, amelyekhez nem szükséges a deriváltak kiszámolása (pl. szimplex módszer)

Hogyan válasszunk a normálegyenletek és az iteratív módszerek között?

- ha nagyon sok ismeretlen paraméter van $\Rightarrow$ a tervmátrix nagyon nagy is lehet
- ekkor gyorsabb lehet valamilyen iteratív módszer
- ha nemlineáris függvényillesztés kell csinálni $\Rightarrow$ iteratív módszer könnyebb lehet, mint a

$$0 = -2 \sum_i \frac{\left[y^{(i)} - h(\mathbf{x}^{(i)}; \mathbf{a})\right]}{\sigma_i^2} \frac{\partial h(\mathbf{x}^{(i)}; \ldots a_k \ldots)}{\partial a_k}$$

megoldása

További fontos kérdések

A paraméterek meghatározásával még nem ért véget a feladat:

- mekkora a meghatározott paraméterek hibája?
- egyáltalán mennyire jó a modell? Hiába kicsi a meghatározott paraméterek hibája, ha rossz a modell, amit használunk