Közelítő szimbolikus számítások

Normák, Sajátértékek

SZTE-TTIK, Számítógépes Optimalizálás Tanszék

Az **x** vektor hosszát a geometriában már értelmeztük, mint a komponensek négyzetösszegének négyzetgyökét. Most ezt a fogalmat általánosítjuk.

Definíció

A $||.|| : \mathbb{R}^n \to \mathbb{R}$ leképezést *vektornormának* mondjuk, ha tetszőleges $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ vektorokra és tetszőleges $\alpha \in \mathbb{R}$ skalárra teljesíti a következő három tulajdonságot:

1. $||\mathbf{x}|| > 0$, ha $\mathbf{x} \neq \mathbf{0}$ ($\mathbf{x}$ nem zérusvektor)
2. $||\alpha\mathbf{x}|| = |\alpha| \cdot ||\mathbf{x}||$ (a konstans szorzó kihozható a normából)
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ (a háromszög-egyenlőtlenség)

Megjegyzés. Hasonlóan értelmezhető *komplex* vektorokra a vektornorma fogalma, csak ekkor a feltételeknek tetszőleges $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ komplex vektorokra és tetszőleges $\alpha \in \mathbb{C}$ komplex skalárra kell teljesülni.

p-normák

Tetszőleges $p >= 1$-re belátható, hogy az

$$||\mathbf{x}||_p = \sqrt[p]{\sum_{1 \le i \le n} |x_i|^p}$$

képlet vektornormát határoz meg. Leggyakrabban használt esetei:

p-normák

- $p = 1$ (egyes norma):

$$||\mathbf{x}||_1 = \sum_{1 \le i \le n} |x_i|$$

- $p = 2$ (kettes vagy euklideszi norma):

$$||\mathbf{x}||_2 = \sqrt{\sum_{1 \le i \le n} |x_i|^2}$$

- $p = \infty$ (végtelen norma):

$$||\mathbf{x}||_\infty = \max_{1 \le i \le n} |x_i|$$

Intuíció

Már korábban is előfordult, hogy a mátrixokat egy számmal jellemeztük, és ez alapján következtettünk tulajdonságaikra, például mátrixok determinánsa $\rightarrow$ létezik-e inverz.
Most a mátrixokhoz fogunk egy újabb számszerű jellemzést megadni, sőt nem is egyet, hanem mindjárt egy egész családot. Ezek lesznek a mátrixnormák.

Definíció

A $||.|| : \mathbb{R}^{n\times n} \to \mathbb{R}$ leképezést *mátrixnormának* mondjuk, ha tetszőleges $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ mátrixokokra és tetszőleges $\alpha \in \mathbb{R}$ skalárra teljesíti a következő négy tulajdonságot:

1. $||\mathbf{A}|| > 0$, ha $\mathbf{A} \neq \mathbf{O}$ ($\mathbf{A}$ nem zérusmátrix)
2. $||\alpha\mathbf{A}|| = |\alpha| \cdot ||\mathbf{A}||$ (a konstans szorzó kihozható a normából)
3. $||\mathbf{A} + \mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||$ (a háromszög-egyenlőtlenség)
4. $||\mathbf{AB}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{B}||$

Megjegyzés. Hasonlóan értelmezhető *komplex* mátrixokra a mátrixnorma fogalma, csak ekkor a feltételeknek tetszőleges $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$ komplex mátrixokra és tetszőleges $\alpha \in \mathbb{C}$ komplex skalárra kell teljesülni.

Példák mátrixnormákra

Legyen $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$.

Néhány gyakran használt mátrixnorma definíciója és a megfelelő Octave függvény hívása:

- 1-es norma (oszlopnorma): $||A||_1 = max_j \sum_{i=1}^{n} |a_{ij}|$

  $$A = \begin{array}{cc} 1 & 2 \\ 3 & -4 \\ 4 & 6 \end{array}$$

  `norm(A,1)`

- végtelen norma (sornorma): $||A||_\infty = max_i \sum_{j=1}^{n} |a_{ij}|$

  $$A = \begin{array}{ccc} 1 & 2 & 3 \\ 3 & -4 & 7 \end{array}$$

  `norm(A,inf)`

- Frobenius norma: $||A||_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2}$

  `norm(A,'fro')`

Sajátérték definíció

Def.: Legyen $\mathbf{A}$ egy négyzetes mátrix. Ha a λ számra és az $\mathbf{x} \neq \mathbf{0}$ vektorra igaz, hogy

$$\mathbf{Ax} = \lambda\mathbf{x}$$

akkor λ az $\mathbf{A}$ mátrix sajátértéke, $\mathbf{x}$ pedig a λ sajátértékhez tartozó sajátvektora.
Megjegyzés. Itt (még valós $\mathbf{A}$ mátrix esetén is!) megengedjük, hogy λ és/vagy $\mathbf{x}$ komplex legyen.

Geometriai értelmezés

- A mátrixok használhatók az n dimenziós tér bizonyos geometriai transzformációinak leírására is.
- A tér egy pontjának az origóból induló, az adott pontban végződő helyvektor felel meg. A transzformációt az adott vektor valamely mátrixszal való balról szorzása végzi el.
- Pl.: Az alábbi **A** mátrix által indukált transzformáció (tükrözés az x tengelyre):
  $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ és $v = [2;1]$ esetén:
  $\mathbf{A}: \quad \mathbf{v} \mapsto \mathbf{A} * \mathbf{v} = \mathbf{v}' = [2, -1]$
- Az **A** mátrix (transzformáció) sajátvektorai azok a vektorok, amelyeket a transzformáció nem forgat el, csak megnyújt. A nyújtás mértéke a sajátvektorhoz tartozó sajátérték.

Példa a sajátértékre, sajátvektorra

Legyen $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$

Egy sajátvektor: $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, a hozzá tartozó sajátérték: $\lambda_1 = 1$

Egy másik sajátvektor: $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, itt a sajátérték: $\lambda_2 = 2$

Sajátérték kiszámítása

Hogyan határozhatók meg a sajátértékek és a sajátvektorok?

- $\mathbf{Ax} = \lambda\mathbf{x} \quad \Longleftrightarrow \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$
- Megfigyelés: a második homogén lineáris egyenletrendszernek csak akkor van nem triviális $\mathbf{x} \neq \mathbf{0}$ megoldása, ha $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- Ha ugyanis az $\mathbf{A}_\lambda = \mathbf{A} - \lambda\mathbf{I}$ determinánsa nem 0, akkor az $\mathbf{A}_\lambda : \quad \mathbf{x} \mapsto \mathbf{A}_\lambda\mathbf{x}$ leképezés kölcsönösen egyértelmű, és a baloldalon csak az $\mathbf{x} = \mathbf{0}$ esetén kaphatunk $\mathbf{0}$ vektort.

Karakterisztikus polinom

- Tehát ha λ sajátérték, akkor $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- Ez a determináns a λ határozatlan polinomja.
- Legyen
  $$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix},$$
  ekkor
  $$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda) = 2 - 3\lambda + \lambda^2$$
  Ezt nevezzük az $\mathbf{A}$ mátrix *karakterisztikus polinomjának*.
- A karakterisztikus polinom gyökei (zérushelyei) a sajátértékek.

Észrevételek:

Tulajdonságok:

- A mátrixnak több sajátértéke lehet (pontosabban ha **A** $n \times n$-es mátrix, akkor n darab, hiszen az n-ed fokú karakterisztikus polinomjának – multiplicitással számolva – pontosan n gyöke van)
- Egy sajátértékhez végtelen sok sajátvektor tartozik (az összes megfelelő irányba mutató)
- Diagonális mátrixok sajátértékei a megfelelő diagonális elemek. Ha például $\mathbf{D} = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix}$, akkor $\det(\mathbf{D} - \lambda \mathbf{I}) = (d_{11} - \lambda)(d_{22} - \lambda)(d_{33} - \lambda)$. Sajátvektorokként vehetők például az adott irányokba mutató egységvektorok.

Definíciók

- A sajátértékek halmaza a *spektrum*. Jele $\lambda(\mathbf{A})$.
  Ha például $\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, akkor $\lambda(\mathbf{A})$ a $\{2, -4, 3\}$ halmaz.
- A sajátértékek abszolút értékének maximuma a *spektrálsugár*. $\rho(\mathbf{A}) = \max\{|\lambda| : \lambda \in \lambda(\mathbf{A})\}$
  A példában $\rho(\mathbf{A}) = 4$.

A sajátérték korlátai

- $\rho(\mathbf{A}) \leq ||\mathbf{A}||$,
  azaz a spektrálsugár mindig kisebb egyenlő, mint a mátrix bármely normája
- Gersgorin körök: Az **A** mátrix minden sajátértéke benne van a
  $$K_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{k=1, k\neq i}^{n} |a_{ik}|\}$$
  körök egyesítésében.

Gersgorin magyarázat

- minden sorhoz egy kör tartozik
- a körök középpontja nálunk az x tengelyen van, mivel valós mátrixokkal foglalkozunk
- a körök sugara pedig a főátlón kívüli elemek abszolútértékösszege

Példa: $\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & -2 & 0 \\ 1 & 0 & 5 \end{bmatrix}$

- Első sorhoz tartozó kör: 1 középpont és 2 sugár
- Második sor: -2 középpont 2 sugár
- Harmadik sor: 5 középpont 1 sugár

Octave függvények a sajátértékek meghatározására

Legyen **A** tetszőleges $n \times n$ dimenziós mátrix. Ekkor

- `D = eig(A)`
  Az n dimenziós **D** oszlopvektor elemei a sajátértékek.
- `[V,D] = eig(A)`
  Az $n \times n$ dimenziós **V** mátrix oszlopai a normalizált sajátvektorok, a **D** diagonális mátrix főátlóbeli elemei pedig a megfelelő sajátértékek.
- `D = eigs(A,k)` (alapértelmezés: $k = 6$)
  A k dimenziós **D** oszlopvektor a k maximális abszolút értékű sajátértéket tartalmazza.
- `[V,D] = eigs(A,k)`
  Az $n \times k$ dimenziós **V** mátrix oszlopai a normalizált sajátvektorok, a $k \times k$ dimenziós **D** diagonális mátrix főátlójában pedig a k maximális abszolút értékű sajátérték áll.

Hatványmódszer

- Föltételezzük, hogy az $\mathbf{A}$ mátrix sajátértékeire $|\lambda_1| > |\lambda_i|, \quad 2 \leq i \leq n$ teljesül.
- Közelítsük a legnagyobb abszolút értékű λ_1 sajátértéket
- és egy hozzá tartozó $\mathbf{x}_1$ sajátvektort.
- Ez iterációs módszer: *közelítő* megoldást kapunk λ_1-re és $\mathbf{x}_1$-re.
- Kell egy *megfelelő* $\mathbf{x}^{(0)}$ kezdővektor.

Az iteráció képlete ($\mathbf{x}^{(0)}$-ból indulva):

$$\mathbf{y}^{(k)} = \mathbf{A}\mathbf{x}^{(k)}$$

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{y}^{(k)}}{||\mathbf{y}^{(k)}||}$$

Itt $||\mathbf{y}^{(k)}||$ az $\mathbf{y}^{(k)}$ vektor normáját jelöli.

Főkomponens analízis

Főkoponens analízis (Principal Component Analysis - PCA):

- Nagy méretű, sok dimenziós adatok esetén szükség lehet a redundancia megszüntetésére, tömörebb tárolásra.
- Szükség lehet arra is, hogy megállapítsuk egy nagy adathalmazban, hogy mely paraméterek okozzák az adatok varianciájának nagy részét.
- Az adatok ábrázolásához azok átalakítására lehet szükség

Ezen feladatok egyik megoldási módszere a Főkomponens analízis, amely egy statisztikai eljárás, de a mesterséges intelligenciában, a felügyelet nélküli tanulásban is az egyik legalapvetőbb módszer.

Példa a PCA-ra

- Tároljunk adatokat autókról. Mekkora a csúcssebességük, fogyasztásuk, hengerűrtartalmuk, milyen a színük, hány ajtósak stb. Az egyszerűség kedvéért legyen minden adat numerikus, és tegyük őket egy X mátrixba, melynek sorai felelnek meg az egyes autóknak, az oszlopok pedig a paramétereknek.
- A leghatékonyabb adatreprezentáció eléréshez az X'*X mátrix (kovariancia mátrix) **sajátvektorainak** meghatározására van szükség. A sajátértékek aránya pedig megmondja, hogy mennyire hatásos a tömörítésünk.

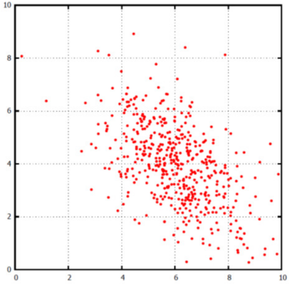
Gyakorló feladat

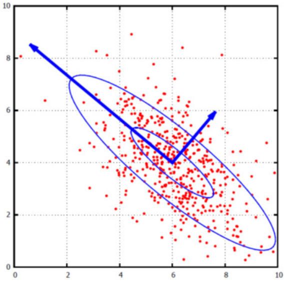
A mellékelt .mat fájlba Octave változókat mentettünk. Benne található egy változó (mátrix), amely emberek magasságát (cm) és tömegét tartalmazza (kg). Forrás: SOCR Data Dinov 020108 HeightsWeights
Az A mátrix a nyers adat, ezt először centralizálni kell (kivonni minden dimenzióból az átlagot)

- Ábrázoljuk az adatpontokat (sorokat) a plot utasítás segítségével. Rajzoljuk ki a data'*data mátrix sajátvektorait ugyanezen azábrán. Mit figyelhetünk meg?
- Tömörítsük az adatot egy dimenzióba! (Olvassunk utána, hogyan működik a PCA. A tömörítés a legnagyobb abszolútértékű sajátvektor irányára történő vetítéssel adható meg). Ábrázoljuk a megfelelő irányú egyenesre vetített pontokat!

1. Gyakorló feladat - 1 hint

Amennyiben jól sikerül a feladat, akkor ehhez hasonló képeket kell kapni (nem pont ugyanezeket):





2. Gyakorló feladat

- Implementáld a hatványmódszert egy olyan függvénnyel, amely 3x3-as mátrixokon működik és minden iterációban 3D-ben kirajzolja az aktuális x^k vektort. (3D-s plotokat a plot3 függvénnyel és a geometry csomaggal tudsz készíteni)