Mašinsko učenje - Matematički podsetnik

Nemanja Mićović, Matematički fakultet

O kursu

Bodovanje

Predispitne obaveze

- 10p - teorijski kolokvijum u kolokvijumskoj nedelji
- 35p - projekat

Završni ispit

- 35p - Teorijski deo (minimum 30%)
- 20p - Praktični deo (minimum 30%)

Projekat

- Vrlo važan deo kursa
- Potrebno posvetiti mu dosta pažnje
- Tema se samoinicijativno predlaže ili bira sa spiska raspoloživih tema
- Spisak tema će biti dostupan nakon kolokvijumske nedelje
- Teme predložene pre kolokvijumske nedelje nose dodatnih 5p
- Projekat se može braniti nezavisno od roka u kojem se polaže završni ispit
- Termini odbrana će biti istaknuti na strani sa obaveštenjima

Kratak matematički podsetnik

Linearna algebra

Vektorski prostor

- Algebarska struktura sa jednom binarnom i jednom spoljnom operacijom $(V, +, \cdot)$
- Binarnu operaciju često označavamo sa $+$
- Vektorski prostor V je *zatvoren* u odnosu na sabiranje:
  - $u \in V$
  - $v \in V$
  - $u + v \in V$

Vektorski prostor

- Spoljnu operaciju obično označavamo sa $\cdot$
  - $u \in V$
  - $\alpha \in K$
  - $\alpha \cdot u \in V$

Vektorski prostor - primeri

- Geometrijski vektori
- Polinomi
- Prostor svih kvadratnih matrica
- Koordinatni prostor R^n

Sopstveni vektor matrice

- Neka je A kvadratna matrica, odnosno $A \in R^{n \times n}$
- Neka je $x \in R^n$, $x \neq 0$
- Ako važi $Ax = \lambda x$ onda je x **desni sopstveni vektor** (right eigenvector)
- Ako važi $xA = \lambda x$ onda je x **levi sopstveni vektor** (left eigenvector)
- λ se naziva **sopstvena vrednost** (eigenvalue)

Sopstveni vektor matrice

$$Ax = \lambda x$$

$$Ax - \lambda x = 0$$

$$(A - \lambda I)x = 0$$

$$det(A - \lambda I) = 0$$

Sopstveni vektor matrice - primer

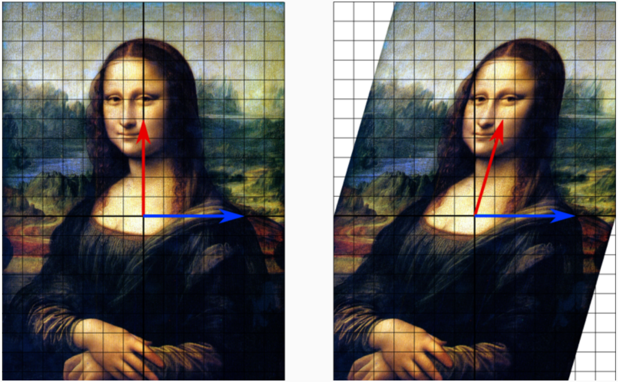
$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

- Odnosno:
  - $\lambda_1 = 1$
  - $\lambda_2 = 3$

Sopstveni vektori

- Vektori koji u transformaciji određenoj matricom A ne menjaju svoj pravac
- Mogu menjati intenzitet i smer



Sopstveni vektori

Linearna zavisnost
$v_1, v_2, ..., v_k$ su *linearno zavisni* vektori ako postoje skalari $\alpha_1, \alpha_2, ..., \alpha_k$ (svi nenula) takvi da je:

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_k v_k = 0$$

Linearna nezavisnost

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_k v_k = 0 \Rightarrow \alpha_1 = \alpha_2 = ... = \alpha_k$$

Sopstveni vektori

Teorema
Sopstveni vektori koji odgovaraju različitim sopstvenim vrednostima su *linearno nezavisni*.

Teorema
Svakoj sopstvenoj vrednosti λ_i odgovara sopstveni podprostor V_i, gde je V_i skup svih vektora x za koje važi $Ax = \lambda_i x$

- $dim(V_i)$ označava dimenziju vektorskog podprostora V_i

Sopstveni vektori

Teorema
Ako je $\sum_i dim(V_i) = dim(A) = n$, A se može svesti na dijagonalnu matricu oblika:

$$D = \left[TODO\right]$$

Odnosno, $AX = DX$ gde je X matrica svih sopstvenih vektora matrice A po kolonama.

Teorema
Ako je A realna simetrična matrica važi $X^{-1} = X^T$, odnosno matrica sopstvenih vektora je **ortogonalna** ($A \cdot A^T = I$ tj $A^{-1} = A^T$). Uvek ju je moguće svesti na **dijagonalnu**.

Definitnost

Pozitivna semidifinitnost
Za kvadratnu matricu $A \in R^{n \times n}$ kažemo da je **pozitivno semidefinitna** ako važi $x^T Ax \geq 0$ za sve $x \in R^n$, $x \neq 0$

Pozitivna definitnost
Za kvadratnu matricu $A \in R^{n \times n}$ kažemo da je **pozitivno definitna** ako važi $x^T Ax > 0$ za sve $x \in R^n$, $x \neq 0$

Definitnost

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \in R^{2\times 2}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$= \begin{bmatrix} x_1 - 2x_2 & 2x_1 + x_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Definitnost

$$= x_1(x_1 - 2x_2) + x_2(2x_1 + x_2)$$
$$= x_1^2 - 2x_1x_2 + 2x_1x_2 + x_2^2$$
$$= x_1^2 + x_2^2 \geq 0$$

- A je **pozitivno definitna**

Definitnost

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$x^T B x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$= \begin{bmatrix} x_1 + x_2 & x_1 + x_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

Definitnost

$$= x_1(x_1 + x_2) + x_2(x_1 + x_2)$$

$$= (x_1 + x_2)^2 \geq 0$$

- Ako je $x_2 = -x_1$ onda je $= 0$ pa je B **pozitivno semidefinitna**

Ispitivanje definitnosti

Teorema (Silvesterov kriterijum)
A je **pozitivno definitna** *akko* je determinanta svih kvadratnih podmatrica koje uključuju elemente a_{11} pozitivna.

Dijagonalno dominantna matrica
Matrica A je dijagonalno dominantna ako važi:

$$|a_{ii}| \geq \sum_{i \neq j} |a_{ij}|, \forall i$$

Teorema
Ako je A dijagonalno dominantna matrica i dijagonalni elementi su nenegativni onda je A **pozitivno semidefinitna**.

Ispitivanje definitnosti

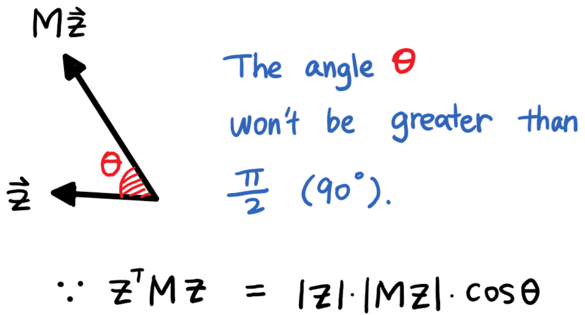
Teorema
Kvadratna simetrična matrica je **pozitivno definitna** *akko* ima **pozitivne sopstvene vrednosti**.

Dokaz (dodati kasnije)

Teorema (Čoleski dekompozicija)
Kvadratna matrica A je **pozitivno definitna** *akko* postoji gornjetrougaona matrica L takva da je $A = L \cdot L^T$.

Smisao definitnosti

- Ukoliko je hesijan funkcije PSD onda je funkcija konveksna
- Ukoliko se vektor x preslika sa matricom koja je PSD, novi smer vektora x će se promeniti za manje od $\frac{\pi}{2}$.



Norma

Neka je $X \subset R^n$. Preslikavanje $\| \cdot \| : X \rightarrow R^+$ za koje važi:

- $\|x\| = 0$ *akko* $x = 0$
- $\|\lambda x\| = |\lambda| \cdot \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$

nazivamo **norma**.

- Intuitivno, norma se povezuje sa intenzitetom vektora (dužinom)

P norme

P norma

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

ℓ_1 norma (Menhetn rastojanje)

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

ℓ_2 norma

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

P norme (ekvivalentnost P normi)

Za svake dve *P norme* važi da postoji $0 < c_1 < c_2$ tako da važi:

$$c_1 \|x\|_a \leq \|x\|_b \leq c_2 \cdot \|x\|_a$$

- Ako ostvarimo konvergenciju u jednoj p-normi, onda imamo tu konvergenciju i u bilo kojoj drugoj p-normi

Normiranje

- Vektor $x \in X$ je **normiran** ako važi $\|x\| = 1$
- Normiranje:
$$\frac{x}{\|x\|}$$

Norme

- Regularizacija je vid popravljanja uslovnosti postavljenog problema
- Norme primenu pronalaze u zadacima regularizacije

Skalarni proizvod

Neka je:

- $X \subset R^n$
- $\langle\rangle : X \times X \to R^+$, $\alpha \in R$
- $+$ operacija sabiranja vektora, $\cdot$ množenje skalarom

Ako važi:

- $\langle a, b\rangle = \langle b, a\rangle$
- $\langle a, b + c\rangle = \langle a, b\rangle + \langle a, c\rangle$
- $\langle a, \alpha b\rangle = \alpha\langle a, b\rangle$
- $\langle a, a\rangle \geq 0$
- $\langle a, a\rangle = 0$ **akko** $a = 0$

Onda $\langle\rangle$ nazivamo **skalarni proizvod**.

Skalarni proizvod - ugao između vektora

- Ugao γ između vektora a i b je u vezi sa skalarnim proizvodom:

$$\langle a, b \rangle = \|a\| \cdot \|b\| \cdot cos(\gamma)$$

- Ako je $\|a\| = 1$ i $\|b\| = 1$ onda:

$$\langle a, a \rangle = cos(\gamma)$$

$$\langle a, a \rangle \in [-1, 1]$$

Kernel

- Jezgro operativnog sistema. . .

Kernel

Kernel

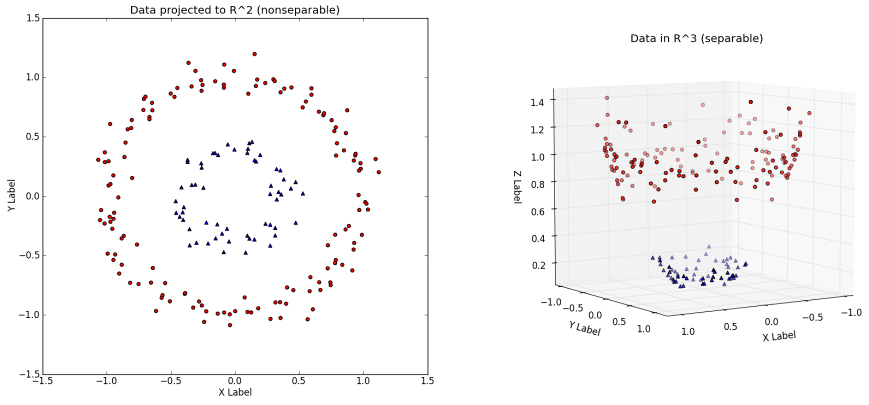
Ako je:

- X neprazan skup $\subset R^n$
- $K : X \times X \rightarrow R$
- Ako je matrica čiji je element $K(x_i, x_j)$ pozitivno semidefinitna za svako $x_1, x_2, ..., x_n \in X$ i $\forall n \in N$

K je pozitivno semidefinitni kernel

Kernel trik

- Intuitivno, kernel predstavlja meru sličnosti elemenata iz skupa X.



Česti kerneli

Polinomijalni kernel:

$$K(x, y) = (\alpha x^T y + c)^d$$

Gausov kernel:

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

Linearni kernel:

$$K(x, y) = x^T y + c$$

Matematička analiza

Izvod funkcije

Izvod funkcije
Za funkciju $f : R \rightarrow R$ izraz:

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

nazivamo **izvod** funkcije f.

Smisao izvoda

- Koeficijent pravca tangente u zadatoj tački

Izvod funkcije

Neka je $f(x) = 2x^2 - x$

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{2(x^2 + 2x\Delta x + \Delta x^2) - x - \Delta x - (2x^2 - x)}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{4x\Delta x + 2\Delta x^2 - \Delta x}{\Delta x} =$$

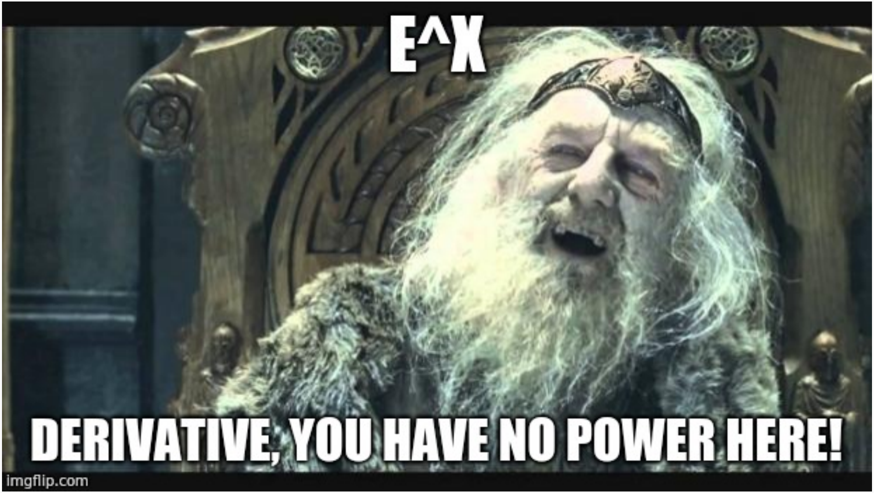
$$= \lim_{\Delta x \to 0} 4x - 1 + 2\Delta x$$

$$f'(x) = 4x - 1$$

Izvod funkcije - poznati rezultati

- $(x^n)' = n \cdot x^{n-1}$
- $(sinx)' = cosx$
- $(cosx)' = -sinx$
- $(e^x)' = e^x$
- $(f + g)' = f' + g'$
- $(f \cdot g)' = f' \cdot g + f \cdot g'$
- $\ldots$

Izvod funkcije



Izvod funkcije

- Od suštinskog značaja kod gradijentnih metoda optimizacije
- Gradijentne metode se koriste pri obučavanju modela mašinskog učenja

Parcijalni izvodi

- Za funkcije više promenljivih ($f : R^n \to R$) uvodimo pojam parcijalnog izvoda

$$\frac{\partial f(x_1, ..., x_n)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_2, ..., x_{i-1}, x_i + h, x_{i+1}, ..., x_n) - f(x_1, ..., x_n))}{h}$$

- Primer:

$$f(x, y) = e^x + y$$

$$f'_x = \frac{\partial f(x, y)}{\partial x} = e^x$$

$$f'_y = \frac{\partial f(x, y)}{\partial y} = 1$$

Parcijalni izvodi

Izvod kompozicije:

$$(g \circ f)(x) = g(f(x))$$

$$\frac{\partial (g \circ f)(x_1, ..., x_n)}{\partial x_i} = \frac{\partial (g \circ f)(x_1, ..., x_n)}{\partial x_i} \cdot \frac{\partial f(x_1, ..., x_n)}{\partial x_i}$$

Parcijalni izvodi - primer

$$f(x, y) = e^{x^2+siny}$$

$$f'_x(x, y) = e^{x^2+siny} \cdot 2x = 2xe^{x^2+siny}$$

$$f'_y(x, y) = e^{x^2+siny} \cdot (cosy) = cosy \cdot e^{x^2+siny}$$

Gradijent

- Vektor svih parcijalnih izvoda funkcije
- $\nabla f = [2xe^{x^2+siny}, cosy \cdot e^{x^2+siny}]$

Hesijan

- Matrica drugih parcijalnih izvoda funkcije

$$\nabla^2 f(m_{ij}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial y} \end{bmatrix}$$

Hesijan

Neka je:

$$f(x, y) = e^x + y$$

Tada:

$$\frac{\partial f}{\partial x} = e^x$$

$$\frac{\partial f}{\partial y} = 1$$

Tako da:

$$\frac{\partial^2 f}{\partial x \partial \partial x} = e^x \quad \frac{\partial^2 f}{\partial x \partial y} = 0$$

$$\frac{\partial^2 f}{\partial y \partial \partial x} = 0 \quad \frac{\partial^2 f}{\partial y \partial \partial y} = 0$$

Hesijan

$$f(x, y) = e^x + y$$

$$\nabla^2 f = \begin{bmatrix} e^x & 0 \\ 0 & 0 \end{bmatrix}$$

Konveksnost

- Funkcija f je konveksna ako važi:

$$f(y) \geq f(x) + \nabla f(x)^T \cdot (y - x)$$

- Još neki uslovi:
  - Zbir konveksnih funkcija je konveksna funkcija
  - Ako je f konveksna funkcija onda je $f(Ax + b)$ takođe konveksna
  - $max\{f_1, ..., f_n\}$ je takođe konveksna funkcija

Jaka konveksnost

- Funkcija koja je diferencijabilna u tački x je jako konveksna ukoliko za svako y iz okoline tačke x važi:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|x - y\|$$

- Jaka konveksnost funkcija je izuzetno poželjno svojstvo u optimizaciji
- Gradijentni optimizacioni algoritmi će uglavnom imati bolje performanse ukoliko je funkcija jako konveksna

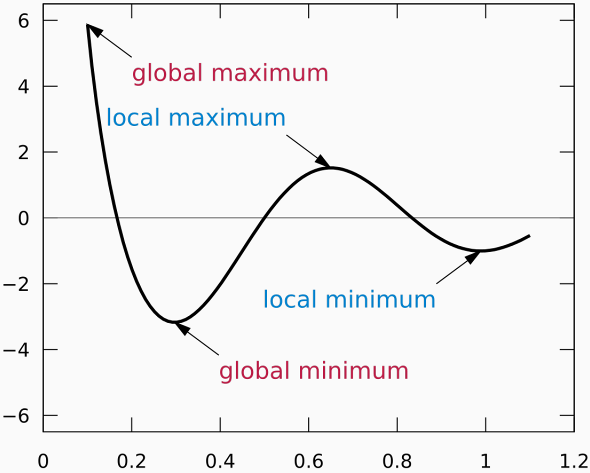
Jaka konveksnost

Teorema
Dva puta diferencijabilna funkcija f je jako konveksna u tački x ako je $\nabla^2 f(x) - mI$ pozitivno semidefinitna za neko m > 0.

Kako u praksi proveriti uslove

- Po definiciji
- Proverom uslova koji važe za hesijan
- Pokazivanje da je funkcija dobijena operacijama koje čuvaju konveksnost

Lokalni optimumi



Lokalni optimumi

Teorema
Ako je x_0 optimum funkcije f i f je diferencijabilna u x_0 tada važi
$\nabla f(x_0) = 0$

- Ako je $f(x_0)$ minimum, onda je funkcija konveksna u tački x_0
- Ako je $f(x_0)$ minimum, onda je funkcija konkavna u tački x_0

Pitanja?