

**WARNING: Math is gonna get methy. So be warned, this doc is gonna meth you up real bad, readers be warned!**

# ITMA jegyzet

Magdi néni előadásai/gyakvideói - DIMAT2 (<https://www.youtube.com/playlist?list=PLhOYgSrY6RXLH0SzTmhm81pdNMeA3J6tv>)

- Csaba: i wanna die, and i will take u with me :D
- Ákos: not if i take you first
- Csaba: hmmmm

## ÁLTALÁNOS CUCCOK, BEVEZETŐ

### 1: alapfogalmak és műveletek

- „Alaki” műveletek:
  - alak megváltoztatása (átméretezés), transzponálás
  - sorok cseréje, oszlopok cseréje
  - kiválasztás, „szeletelés” (vágás, slicing), összeillesztés
- Számmal szorzás megegyezik a matematikai művelettel.  
 $M = c * A$  esetén  $M[i,j] = c * A[i,j]$
- Összeadás (kivonás) megegyezik a matematikai művelettel.  
 $M = A + B$  esetén  $M[i,j] = A[i,j] + B[i,j]$
- Szorzás nem egyezik meg a matematikai művelettel.  
 $M = A * B$  esetén  $M[i,j] = A[i,j] * B[i,j]$
- „Reciprok” („osztás”) nem egyezik meg a matematikai művelettel.  
 $M = 1 / B$  esetén  $M[i,j] = 1 / B[i,j]$   
 $M = A / B$  esetén  $M[i,j] = A[i,j] / B[i,j]$
- Hasonlóan egyéb műveletek és függvények, általában nem egyeznek meg a matematikai művelettel.  
 $M = A ** B$  esetén  $M[i,j] = A[i,j] ** B[i,j] : A^B$   
 $M = \exp(A)$  esetén  $M[i,j] = \exp(A[i,j]) : e^A$
- Mi van a „rendes” matematikai műveletekkel?
  - **Szorzás (dot product)**
    - A.dot(B) elvégezhető vektorokon(1D array) vagy mátrixokon(min 2D array)

- Ha vektorokon végezzük el akkor inner productnak hivhatjuk
- Ha mátrixon végezzük akkor mátrix szorzásnak felel meg (matrix multiplication)
  - de a numpy matmul vagy A @ B operáció is ugyan azt csinálja
  - 2D esetében az első mátrix oszlopszámának meg kell egyeznie a második mátrix sorainak számával => így működik:
    - A.shape == (m,n) és B.shape == (n,p) és M.shape == (m,p)
    - $M[i,j] = (A[i,:]) * (B[:,j]).sum()$
- $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$        $B = \begin{bmatrix} 2 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 \end{bmatrix}$        $A.shape = (m,n) = (4,2)$   
 $B.shape (n,p) = (2,4)$   
 $A @ B$   
 $M[1,1] = A[1,:] * B[:,1].sum()$        $M, shape = (4,4)$   
 $M[1,1] = (1,2) \cdot (1+1) = (1,2) \cdot 2$

- **Inverz** (költséges számítás).

- $M = np.linalg.inv(A)$
- $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
- DE Ha van egy diagonális mátrixod (a nem fő átlóba CSAK 0-k vannak) pl Hesse mátrix akkor a tagok reciprokát kell venni
  - $\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$

## 2.:TÖMBÖK

- A kétdimenziós tömbnek két mérete van:
  - 1.: sorok száma
  - 2.: oszlopok száma (ilyen sorrendben).
  - Az  $m \times n$  méretű M mátrix mérete:  $M.shape = (m,n)$
  - 2D tömb szorzása 1D tömbbel:
    - 2D minden sorát!!!
    - skalárisan szorozzuk 1D-vel
- A háromdimenziós tömbnek három mérete van:
  - rétegek száma
  - sorok száma
  - oszlopok száma (ilyen sorrendben).
  - A k rétegben  $m \times n$  méretű márixokat tartalmazó M 3D tömb mérete:  $M.shape = (k,m,n)$
- **3D** tömb szorzása **1D** tömbbel:
  - 3D minden rétegében!!!
  - minden sorát!!!
  - skalárisan szorozzuk 1D-vel

- $(k, m, n)$  és  $(n,)$  szorzás eredménye  $(k, m)$  méretű.
- **3D** tömb szorzása **2D** tömbbel
  - 3D minden rétegét!!!
  - szorozzuk 2D-vel.
  - $(k, m, n)$  és  $(n, p)$  szorzás eredménye  $(k, m, p)$  méretű.
- Ez **3D+** tömb szorzása:
  - When calculating the dot product( $A @ B$ ) of two N-D arrays  $A$  and  $B$  using `np.dot(A, B)`, the result is a sum product over the last axis of  $A$  and the second-to-last axis of  $B$ . Specifically, if  $A$  has shape  $(i, j, k)$  and  $B$  has shape  $(k, l, m)$ , then the resulting array  $M$  will have shape  $(i, j, l, m)$

## BROADCASTING

---

- Broadcasting csodái :D -> python array broadcasting  
(<https://jakevdp.github.io/PythonDataScienceHandbook/02.05-computation-on-arrays-broadcasting.html>)
- **CSAK bal oldalról** lehet bővíteni a kisebb méretű mátrixot. 1-esekkel bővítjük és az 1-eseket tudjuk utána kinyújtani
- **Broadcasting** => dimezió kiterjesztést jelent
- összeadásnál alkalmazható
- Ha valamelyik dimenzió **1**, akkor kiterjeszthető, különben nem lehet
- A nem-létező dimenziók mindig elől állnak, azaz hátulról (jobbról) olvasva kell a méreteknek egyezni.
- FONTOS: (2,) ez egy csak egy skalár (szám) nem vektor

- 3 szabály

- **Rule 1:** If the two arrays differ in their number of dimensions, the shape of the one with fewer dimensions is padded with ones on its leading (left) side.

- skalárral/0D-s mátrixtal való broadcasting

`M.shape = (2, 3) a.shape = (3,) => (1, 3) => (2, 3)`

- **Rule 2:** If the shape of the two arrays does not match in any dimension, the array with shape equal to 1 in that dimension is stretched to match the other shape.

`M.shape = (2, 3) a.shape = (1, 3,) => (2, 3)`

- **Rule 3:** If in any dimension the sizes disagree and neither is equal to 1, an error is raised.

```

M = np.ones((3, 3))
print('a:\n', a)
print('M:\n', M)
print('\n-----\na+M:\n', a+M)

[17]   ✓  0.0s
...   a:
[0 1 2]
M:
[[1. 1. 1.]
 [1. 1. 1.]
 [1. 1. 1.]]

-----
a+M:
[[1. 2. 3.]
 [1. 2. 3.]
 [1. 2. 3.]]

```

- Tehát:

- ha nem megfelelő a dimenziók száma, 1-essel bővíthetünk a bal oldalon
- ha 2 mátrix csak abban térnek el hogy az egyikben adott dimenzióknál 1-es szám van akkor a kettő összeadható, mert azt ki tudjuk nyújtani

`np.arange(3)+5`

$$\begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 5 & 5 & 5 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 5 & 6 & 7 \\ \hline \end{array}$$

`np.ones((3, 3))+np.arange(3)`

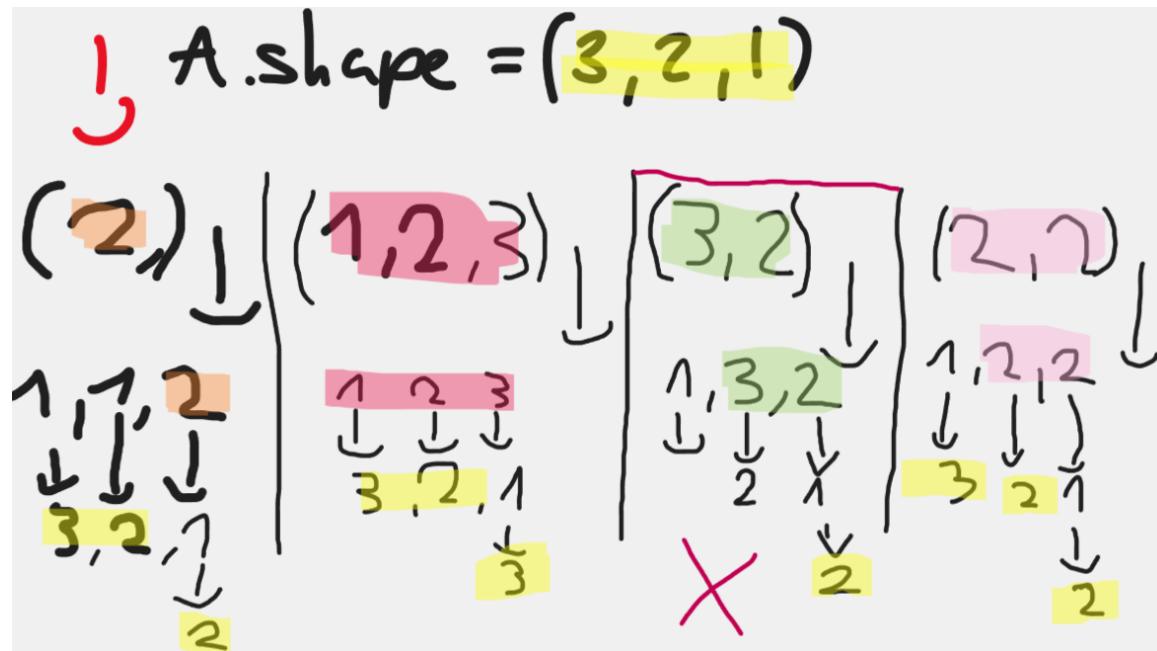
$$\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 0 & 1 & 2 \\ \hline 0 & 1 & 2 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 1 & 2 & 3 \\ \hline 1 & 2 & 3 \\ \hline \end{array}$$

`np.arange(3).reshape((3, 1))+np.arange(3)`

$$\begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 2 & 2 & 2 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 0 & 1 & 2 \\ \hline 0 & 1 & 2 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 1 & 2 & 3 \\ \hline 2 & 3 & 4 \\ \hline \end{array}$$

- Megjegyzés: Ahogy a képen is látszik, minden a kettő mátrix adott dimenzióját lehet nyújtani

- Példa:



## NORMA

- Vektor: 1 dimenziós mátrix
- Vektortér: Vektortér kb. amelynek elemei korlátlanul és egyértelműen **összehatásos** és számmal **szorozhatók**.
- Def:  $\|\cdot\| : V \rightarrow \mathbb{R}_0^+$
- Vektor hosszának számítása:
  - általános alak:  $|a| = \sqrt{\sum_{n=1}^n a_n^2}$
  - vegyük  $a=[2, 2, 3]$  vektort
    - ebben az esetben:  $|a| = \sqrt{2^2 + 2^2 + 3^2} = \sqrt{17}$

- Norma: Általánosított vektorhossz
- Szabályok:**
  - $\|v\| = 0 \iff v = 0$ , ha a vektor hossza 0, maga a vektor is 0
  - $\forall v \in V$  és  $\forall \lambda \in \mathbb{T}$  esetén  $\|\lambda v\| = |\lambda| \cdot \|v\|$  (homogenitás / skálázhatóság)
  - Magyarál:**
    - vessük  $v=[1, 2, 3]$  vektort  $\lambda=2$ -vel
    - $\sqrt{(1 \cdot 2)^2 + (2 \cdot 2)^2 + (3 \cdot 2)^2} = 2 \cdot \sqrt{1^2 + 2^2 + 3^2}$
  - $\forall u \in V$  és  $\forall v \in V$  esetén  $\|u + v\| \leq \|u\| + \|v\|$ . (szubadditivitás / háromszög-egyenlőtlenség)
    - Ha vektorokat összeadjuk és utána vessük a hosszát az kisebb egyenlő azzal hogy először minden vektor hosszát vessük és utána adjuk össze

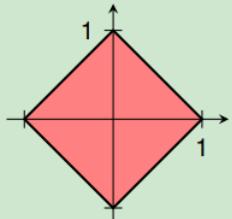
## Hölder-norma (vagy $p$ -norma)

Ha a  $\mathbf{v}$  vektor koordinátái egy rögzített bázisban:  $\mathbf{v} = [x_1 \ x_2 \ \dots \ x_n]^T$  és  $p \geq 1$ , akkor

$$\|\mathbf{v}\|_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{\frac{1}{p}} = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}.$$

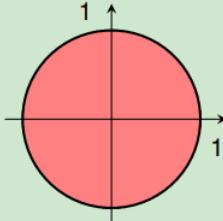
## Egységsugarú körök (környezetek) a 2D valós vektortéren

$$\|\mathbf{v}\|_1 = |x_1| + |x_2|$$



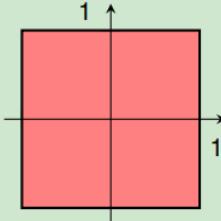
Manhattan-norma

$$\|\mathbf{v}\|_2 = \sqrt{x_1^2 + x_2^2}$$



Euklidészi norma

$$\|\mathbf{v}\|_\infty = \max \{|x_1|; |x_2|\}$$



Csebisev-norma

- Példa:

2)  $\|(6; -2; -3)\|_1$   
 norma által. elc.  $\Rightarrow \sqrt{\sum_{i=1}^3 n_i^2} \rightarrow \|\mathbf{v}\|_2$   
 $\sqrt{6^2 + (-2)^2 + (-3)^2} = \sqrt{36 + 4 + 9} = 7$

Manhattan norma:  $\sum |n_i| \rightarrow \|v\|_1$   
 $6 + 2 + 3 = 11$

Pálosi Ákos

Csebisev norma:  $\max\{|x_1|; |x_2|; \dots; |x_n|\} \rightarrow \|v\|_\infty$   
 $\max\{6; 2; 3\} = 6$

## METRIKA

- Általánosított távolság.
- Két vektor távolsága
  - $v = [v_1, v_2], w = [w_1, w_2]$
  - $d(v; w) = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2}$
- Adott  $H$  halmaz (nem feltétlenül vektortér). A

- $d : H \times H \rightarrow R_0^+$
- függvény **metrika**, ha
  1.  $d(x; y) = 0 \iff x = y$
  2.  $\forall x, y \in H$  esetén  $d(x; y) = d(y; x)$  - **szimmetria**
  3.  $\forall x, y, z \in H$  esetén  $d(x; z) \leq d(x; y) + d(y; z)$ . - **háromszög-egyenlotlenség**

- $V$  minden normájából származtatható metrika  $V \times V$ -n:
  - $d(u; v) = \|u - v\|$ .
- $V \times V$  minden metrikájából származtatható norma  $V$ -n:
  - $\|v\| = d(\theta; v)$ .

### Példa a 3 szabályra

- $v = [1, 2, 3], w = [4, 5, 6], t = [7, 8, 9]$

- 1.  $v=v, d(v,v)=0 =>$ 
  - $d(v, v) = \sqrt{(1-1)^2 + (2-2)^2 + (3-3)^2} = 0$
- 2.  $d(v; w) = d(v; w) =>$ 
  - $d(v, w) : \sqrt{(1-4)^2 + (2-5)^2 + (3-6)^2} = 3\sqrt{3} =$   
 $= d(w, v) : \sqrt{(4-1)^2 + (5-2)^2 + (6-3)^2}$
- 3.  $d(v; t) \leq d(v; w) + d(w; t) =>$ 
  - $d(v; t) : \sqrt{(1-7)^2 + (2-8)^2 + (3-9)^2} = 3\sqrt{3} \leq$   
 $d(v; w) : \sqrt{(1-4)^2 + (2-5)^2 + (3-6)^2} + d(w; t) : \sqrt{(4-7)^2 + (5-8)^2 + (6-9)^2}$   
 $= 3\sqrt{3} + 3\sqrt{3} = 6\sqrt{3}$

- **MAE RMSE/MSE:** becslési hibák megtalálása

$$x = [1,0 \quad 2,0 \quad 3,0 \quad 4,0] \quad y = [1,0 \quad 2,0 \quad 3,0 \quad 4,0 \quad 5,0 \quad 6,0 \quad 7,0 \quad 8,0 \quad 9,0] \\ \hat{x} = [0,9 \quad 2,1 \quad 3,1 \quad 3,9] \quad \hat{y} = [0,9 \quad 2,1 \quad 3,1 \quad 3,9 \quad 5,1 \quad 6,1 \quad 6,9 \quad 7,9 \quad 9,1]$$

Látszólag egyformán jó becslések, pedig

$$d_1(x; \hat{x}) = \|x - \hat{x}\|_1 = 0,4 \quad d_1(y; \hat{y}) = \|y - \hat{y}\|_1 = 0,9 \\ d_2(x; \hat{x}) = \|x - \hat{x}\|_2 = 0,2 \quad d_2(y; \hat{y}) = \|y - \hat{y}\|_2 = 0,3$$

- Megoldás: átlagoljunk

$$\text{MAE} = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n} \quad (\text{Mean Absolute Error}) \quad \text{MAE}(\hat{x}) = 0,1 = \text{MAE}(\hat{y})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \quad (\text{Root Mean Squared Error}) \quad \text{RMSE}(\hat{x}) = 0,1 = \text{RMSE}(\hat{y})$$

$$\text{MSE} = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n} \quad (\text{Mean Squared Error}) \quad \text{MSE}(\hat{x}) = 0,01 = \text{MSE}(\hat{y})$$

- $x$  becslései 0.1 el térnek el ( $1-0.9=0.1$ ) ezeket összeadva jön ki a 0.4

# BÁZIS, VEKTORTÉR

ELTE vektorok és vektor terek

([https://nimbus.elte.hu/~hagi/segedanyag/vektorszamitas\\_felev2/vektor\\_felev2\\_jegyzet2019.pdf](https://nimbus.elte.hu/~hagi/segedanyag/vektorszamitas_felev2/vektor_felev2_jegyzet2019.pdf))

- **lineáris függetlenség:** A lineáris algebrában vektorok egy halmazát lineárisan függetlennek nevezük, ha egyikük sem fejehető ki a többi vektor lineáris kombinációjaként. Ellenkező esetben lineárisan összefüggő vektorokról beszélünk.
- **Síkvektort definiáló szabályok**
  1. Az elemek irányított szakaszok, amelyeket nyíllal szemléltethetünk.
  2. Az elemeknek van nagyságuk és irányuk.
  3. Bármely két elemnek értelmezve van az összege, és ez az összeg halmazbeli, azaz annak egy adott eleme.
  4. Bármely elemnek értelmezve van egy valós számmal való szorzása (a skalárszorosa), és a skalárszoros is a halmaz eleme.
  5. Az elemek között értelmezve van skaláris szorzás. Vigyázat: ez nem ugyanaz, mint az el® bbi pontban említett skalárral való szorzás! A skaláris szorzás eredménye nem a halmaz eleme, hanem egy valós szám

- **Vektortér:** Az összes definiált síkvektor halmaza

**Lineáris kombináció:** Skalárok és vektor elemeinek szorzata

- $\lambda_1 v_1 + \lambda_2 v_2 \dots$
- Az  $a_1, \dots, a_n \in V$  vektorokat a V vektortér **generátorrendszerének** nevezük, ha V minden eleme előáll az  $a_i$  vektorok **lineáris kombinációjaként**.

- Vektortér **bázisának** nevezük a vektortér **lineárisan független generátorrendszerét**.

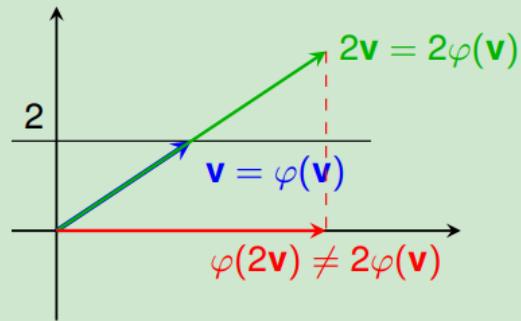
## LINEÁRIS LEKÉPEZÉSEK

- Lineáris leképezés
- $V_1$  és  $V_2$  ugyanazon T test feletti vektorterek.
- A  $\varphi: V_1 \rightarrow V_2$  függvény lineáris leképezés, ha
  - $\forall a, b \in V_1$  esetén  $\varphi(a + b) = \varphi(a) + \varphi(b) \Rightarrow \varphi$  **összegtartó** = két vektor összegének képe a két vektor képének összege
  - $\forall \lambda \in T, \forall a \in V_1$  esetén  $\varphi(\lambda a) = \lambda \varphi(a) \Rightarrow \varphi$  **aránytartó** = egy vektor számszorosának képe a vektor képének ugyanezen számszorosa

## Lineáris transzformációk

- $\varphi$  lineáris leképezés **lineáris transzformáció**, ha  $V_1 = V_2$

- $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , tükrözés  $y = 2$  egyenesre



NEM lineáris transzformáció.

- $\varphi: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ ,  $\varphi(\mathbf{X}) = \mathbf{X} + \mathbf{C}$ , ahol  $\mathbf{C} \neq \mathbf{0}$  adott mátrix.

- $\varphi(\mathbf{X}_1 + \mathbf{X}_2) = (\mathbf{X}_1 + \mathbf{X}_2) + \mathbf{C} \neq (\mathbf{X}_1 + \mathbf{C}) + (\mathbf{X}_2 + \mathbf{C}) = \varphi(\mathbf{X}_1) + \varphi(\mathbf{X}_2)$ ;
- $\varphi(\lambda \mathbf{X}) = (\lambda \mathbf{X}) + \mathbf{C} \neq \lambda (\mathbf{X} + \mathbf{C}) = \lambda \varphi(\mathbf{X})$ .

NEM lineáris transzformáció.

- $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , tükrözés  $y = 2$  egyenesre  
 $\varphi\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$   $\Rightarrow$  NEM lineáris transzformáció.
- Eltolások, azaz a  $\varphi(x) = x + c$ , ( $c \neq 0$ ) típusú függvények NEM lineáris transzformációk; (mint ahogyan a mátrixos példában láttuk).
- A feltétel csak szükséges, de nem elég. A  $\varphi: \mathbb{C} \rightarrow \mathbb{C}$ ,  $\varphi(z) = z^2$  transzformáció esetén  $\varphi(0) = 0^2 = 0$ , de láttuk, hogy nem lineáris.

- $\varphi: \mathbb{C} \rightarrow \mathbb{C}$ ,  $\varphi(z) = \bar{z}$ . Lineáris transzformáció, mert

- $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$ ;
- $\overline{\lambda z} = \bar{\lambda} \bar{z}$ .

- $\varphi: \mathbb{C} \rightarrow \mathbb{C}$ ,  $\varphi(z) = (1 - 2j)z$ . Lineáris transzformáció, mert

- $\varphi(z_1 + z_2) = (1 - 2j)(z_1 + z_2) = (1 - 2j)z_1 + (1 - 2j)z_2 = \varphi(z_1) + \varphi(z_2)$ ;
- $\varphi(\lambda z) = (1 - 2j)(\lambda z) = \lambda(1 - 2j)z = \lambda \varphi(z)$ .

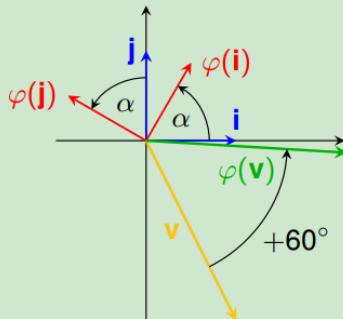
- Ha  $\varphi$  lineáris transzformáció, akkor  $\varphi(\mathbf{0}) = \mathbf{0}$

- eeevvileg: Azért, mert 0-t behelyettesítve ki tudjuk deríteni, hogy az értéke megváltozik-e a transzformáció során. Ha NEM változik, akkor lineáris transzformációról beszélünk

## Transzformáció mátrixa

- Magyarul: a transformációt leíró mátrix

- Origó körüli,  $\alpha$  szögű forgatás (síkban).  $B(\mathbf{i}, \mathbf{j})$



$$\varphi(\mathbf{i}) = \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix} \quad \varphi(\mathbf{j}) = \begin{bmatrix} -\sin \alpha \\ \cos \alpha \end{bmatrix}$$

$$\mathbf{F}_\alpha = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}.$$

Forgassuk el a  $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$  vektort origó körül  $60^\circ$ -kal:

$$\begin{aligned} \varphi(\mathbf{v}) &= \mathbf{F}_{60^\circ} \mathbf{v} = \begin{bmatrix} \cos 60^\circ & -\sin 60^\circ \\ \sin 60^\circ & \cos 60^\circ \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \\ &= \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} + \sqrt{3} \\ \frac{\sqrt{3}}{2} - 1 \end{bmatrix} \approx \begin{bmatrix} 2,232 \\ -0,134 \end{bmatrix}. \end{aligned}$$

## Összetett transzformáció

- Adottak
  - $\varphi_1, \varphi_2 : V \rightarrow V$  lineáris transzformációk
  - az (ugyanabban a bázisban felírt)  $A_1, A_2$  mátrixokkal írhatjuk le a  $\varphi_1, \varphi_2$ -t
  - $V$  vektorain végezzük el előbb a  $\varphi_1$ , majd a  $\varphi_2$  transzformációt, azaz alkalmazzuk a  $\varphi_2 \circ \varphi_1$  kompozíciójukat
- Fentiekből jön, hogy:
  - $(\varphi_2 \circ \varphi_1)(v) = \varphi_2(\varphi_1(v)) = \varphi_2(A_1 v) = A_2(A_1 v) = (A_2 A_1)v$
  - Magyarul:** transzformáljuk a  $v$ -t és ennek az eredményét is transzformáljuk
  - Tehát  $(\varphi_2 \circ \varphi_1)$  mátrixa  $A_2 A_1$

Nagyon ügyeljünk a **sorrendre**: a mátrixok szorzása **NEM kommutatív**. Az előbb alkalmazandó transzformáció mátrixa áll hátrébb a szorzatban

- $\varphi : \mathbb{C} \rightarrow \mathbb{C}$ , minden komplex számot megszoroz  $(1 - 2j)$ -vel, majd konjugálja.

$$\varphi_1 : \mathbb{C} \rightarrow \mathbb{C}, \quad \varphi_1(z) = (1 - 2j)z, \quad \mathbf{A}_1 = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

$$\varphi_2 : \mathbb{C} \rightarrow \mathbb{C}, \quad \varphi_2(z) = \bar{z}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Mivel  $\varphi = \varphi_2 \circ \varphi_1$ , ezért  $\varphi$  mátrixa:

$$\mathbf{A} = \mathbf{A}_2 \mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}.$$

## KÉP ÉS MAGTÉR

- megjegyzés:  
egy  $R^{4x20}$  as transzformációs mátrix esetében a 20 dimenziós vektorokat  
**legfeljebb** 4 dimenzióva képezzük

### Képtér

#### DEF:

- $\varphi : V \rightarrow V$  lineáris transzformáció
- értékkészlete  $\varphi(V) : \varphi$ 
  - $\varphi(V) = \{\varphi(v) | v \in V\} = \{Av | v \in T^n\} = \langle a_1, \dots, a_n \rangle.$

- Tehát:
  - $R^n$  lineáris leképezésének az értékkészlete =  $R_\varphi$ 
    - az értelmezési tartományon vett értékkészlet
  - Transzformáció eredménye a képtér
  - $\varphi$  értékkészlete **altér**  $V$ -ben ( $V_2$ -ben), ezért
    - képtér**-nek nevezzük
    - jele**:  $\text{Im}(\varphi)$ .
  - $\varrho(A) = \dim \text{Im}(\varphi)$

- 30°-os egyenesre vetítés:  $\mathbf{P}_{30^\circ} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix} \Rightarrow \varrho(\mathbf{P}_{30^\circ}) = 1;$   
 $\text{Im}(\varphi) = \left\{ \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} t \mid t \in \mathbb{R} \right\} \Rightarrow \dim \text{Im}(\varphi) = 1.$
- $\varphi : \mathbb{R}_3[x] \rightarrow \mathbb{R}_3[x] \quad \varphi(p) = p' \quad (\text{polinom deriválás})$   
 $\mathbf{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \varrho(\mathbf{D}) = 3;$   
 $\text{Im}(\varphi) = \mathbb{R}_2[x] \Rightarrow \dim \text{Im}(\varphi) = 3.$

- az oszlopvektorainak(x vektorokok) vett lineáris kombinációja:
  - $A = [a_1, \dots, a_n]$
  - $x = [x_1, \dots, x_n]$
  - $Ax = x_1a_1 + x_2a_2 + \dots + x_na_n$

### Magtér

- tengelymetszetek, zérushelyek halmaza
  - olyan  $x$ -ek az értelmezési tartományból amikre a  $\varphi$  az  $x$  helyen **null vektor**

- $\text{Ker}(\varphi) = \{x \in R^n | \varphi(x) = 0\}$
- olyan  $n$  dimenziós vektorokból áll amik 0-ra vetítik a mátrixot (ún. **Null tere** =  $N(A)$ )
  - homogén lineáris egyenlet rendszer megoldásai

Általános példa:

- Adott az  $A \in R^{kxn}$  mátrix
  - képtere:  $\text{Im } A = \{Ax | x \in R^n\}$
  - magtere:  $\text{Ker } A = \{x \in R^n | Ax = 0\}$

Példa mátrixal

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 1 & 2 & 1 & -1 & 0 \\ -1 & -2 & 4 & 6 & 5 \\ 1 & 2 & 2 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 1 & 2 \\ 0 & 0 & -2 & -2 & -2 \\ 0 & 0 & 7 & 7 & 7 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & -2 & -1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow x_1 = -2r + 2s + t$$

$$\underline{s}_2 - \underline{s}_1$$

$$\underline{s}_3 + \underline{s}_1$$

$$\underline{s}_4 - \underline{s}_1$$

$$\underline{s}'_2 = \underline{s}_2 \cdot (-2)$$

$$\underline{s}'_3 = 7\underline{s}'_2$$

$$\underline{s}'_4 = \underline{s}'_2$$

$$\underline{s}'_1 = 3\underline{s}'_2$$

$$\text{g}(A) = 2 = \dim(\text{g}(A))$$

Lépési sorrend: 1. 3. oszlopban.

$\text{g}(A)$  egy bázisa:  $\underline{a}_1, \underline{a}_2$

$$\underline{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \underline{a}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

$$\underline{a}_2 = 2 \cdot \underline{a}_1$$

$$\underline{a}_4 = -2\underline{a}_1 + \underline{a}_3$$

$$\underline{a}_5 = -\underline{a}_1 + \underline{a}_3$$

$$A\underline{x} = \underline{0} \quad \text{más:}$$

$$\begin{aligned} x_2 &= r \in \mathbb{R} \\ x_4 &= s \in \mathbb{R} \quad \text{teh.} \\ x_5 &= t \in \mathbb{R} \end{aligned}$$

$$\begin{aligned} x_1 &= -2r + 2s + t \\ x_3 &= -s - t \end{aligned}$$

Ker A elemei:

$$\begin{bmatrix} -2r + 2s + t \\ r \\ -s - t \\ s \\ t \end{bmatrix} = r \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Ker A egy bázisa:

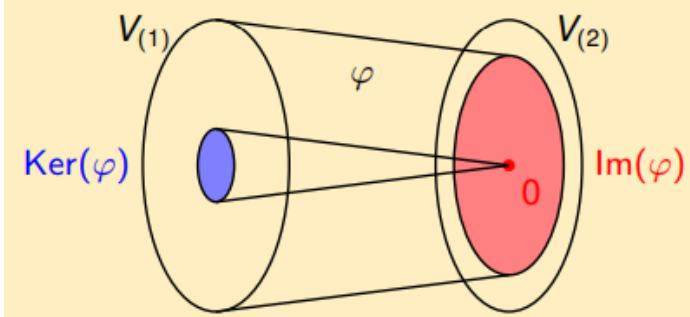
$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\dim \text{Ker } A = 3 = \text{a szabad bar. száma.}$$

## Dimenzió-tétel

- Ha  $\varphi : V(1) \rightarrow V(2)$  lineáris transzformáció (leképezés), akkor:

$$\dim \text{Im}(\varphi) + \dim \text{Ker}(\varphi) = \dim V_{(1)}.$$

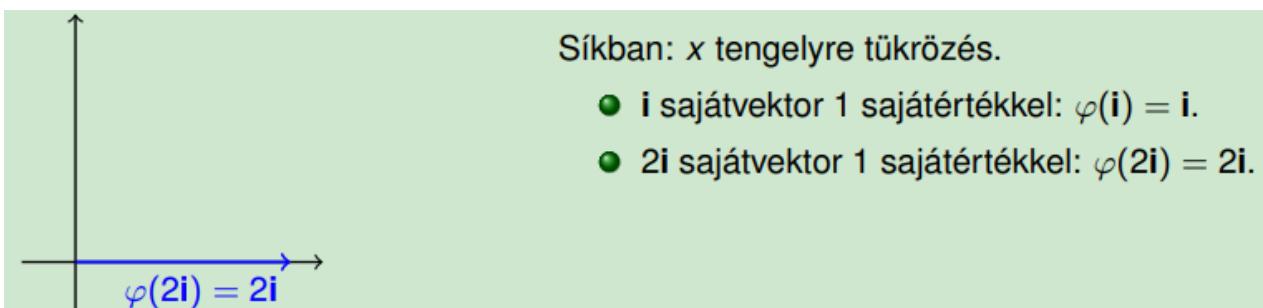


## Sajátérték és sajátvektor

- BME-s papír -> definíciók, példák, feladatmegoldások  
([https://math.bme.hu/algebra/a2/2009/5\\_Sajatertek.000.pdf](https://math.bme.hu/algebra/a2/2009/5_Sajatertek.000.pdf))
- $A\varphi : V \rightarrow V$  lineáris transzformáció:
  - sajátvektora:**  $s \in V \setminus \{0\}$  vektor
    - értékkészlete: saját maga kivéve a 0 vektor
    - saját vektora :  $v$ , ha  $Av = \lambda v$
  - sajátértéke:**  $\lambda \in T$ , ha  $\varphi(s) = \lambda s$ 
    - $\lambda \in R$  szám esetén  $\rightarrow \lambda$ -t a  $v$ -hez tartozó saját értéknek nevezzük

- Magyarul:
  - Ha  $A$  mátrixot beszorozzuk egy  $\lambda$  számmal és az eredmény párhuzamos az  $A$  mátrix és  $v$  vektor szorzatával.
  - Ekkor a  $v$  vektort az  $A$  mátrixnak a **sajátvektorának**, a  $\lambda$ -t pedig  $v$  vektorhoz tartozó **sajátértékének** nevezzük

- Megjegyzés:
  - Minden  $\varphi$  lin. transf. esetén  $\varphi(0) = 0 = \lambda \cdot 0$ .
  - Illetve 0 vektor esetén minden párhuzamos lesz az eredmény, ezért excludejük a sajátvektorok halmazából



$$A \underline{v} \parallel \underline{v} \quad (\text{Ha } \underline{v} = \underline{0}, \text{ akkor } A\underline{v} = \underline{0} \parallel \underline{0} = \underline{v})$$

Def:  $A \underline{v} \neq \underline{0}$   $\underline{v} \in \mathbb{R}^n$  vertélt az  $A$  ( $n \times n$ )-es mátrix sajátbelterületén

$$A \underline{v} = \lambda \underline{v}$$

omely  $\lambda \in \mathbb{R}$  számnal. Ekkor  $\lambda$ -t a  $\underline{v}$ -hez tartozó sajátértéket nevezik, ha

## Saját-altér

- DEF: a négyzetes  $A$  mátrix  $\lambda$  **sajátérték**hez tartozó **sajátvektorai** és a **nullvektor** alkotta alteret a  $\lambda$  sajátértékhez tartozó **sajátaltérnek** nevezük

- $\varphi(s_1 + s_2) = \varphi(s_1) + \varphi(s_2) = \lambda s_1 + \lambda s_2 = \lambda(s_1 + s_2)$
- $\varphi(\mu s) = \mu \varphi(s) = \mu(\lambda s) = \lambda(\mu s)$
- sajátvektor definiálásánál 0-t kivettük a vektorterünk ből és nem vettük figyelembe itt a **0 is tagja a vektorterünknek**

## Sajátérték kiszámítása

- Az egyetlen változó amit nem ismerünk az a  $\lambda$  vagyis, a sajátérték
- Legyen  $s \in V$  a  $\varphi : V \rightarrow V$  lineáris transzformáció
  - sajátvektora:  $\lambda \in T$
  - sajátértékkel:  $A \in T^{n \times n}$
  - a  $\varphi$  mátrixa egy rögzített bázisban,  $s \in T$
  - $n$  pedig  $s$  koordinátái ugyanabban a bázisban

$$\begin{aligned}\varphi(s) &= \lambda s, \\ As &= \lambda s, \\ As - \lambda s &= 0, \\ (A - \lambda) s &= 0, \\ (A - \lambda E) s &= 0\end{aligned}$$

- Az  $E$  egységmátrixra azért van szükség mert  $\lambda s$  skalár, és azt nem tudjuk kivonni a mátrixból (technikailag am de, de nnna ez így jó és kész), viszont mátrix minusz mátrix már értelmezhető
- Homogén lineáris egyenletrendszer** kaptunk, amelynek  $s = 0$  **mindig** megoldása.
- Nem-triviális** megoldást úgy kaphatunk, ha a lineáris egyenletrendszernek nem egyértelmű a megoldása. Ennek létezését (pl. Cramer-szabály alapján) úgy biztosíthatjuk, ha

$$\det(A - \lambda E) = 0.$$

- Ez a transzformáció **sajátérték-egyenlete**, vagy a **sajátértékek karakterisztikus egyenlete**

Számítsuk ki annak a transzformációnak a sajátértékeit és sajátvektorait, amelynek mátrixa  $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$

$$\det(\mathbf{A} - \lambda \mathbf{E}) = \det \left( \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right)$$

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = (1-\lambda)(-2-\lambda) - 4$$

$$\lambda^2 + \lambda - 6 = 0 \Rightarrow \lambda_1 = 2, \quad \lambda_2 = -3$$

•  $\lambda_1 = 2$

$$\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow \begin{cases} x + 2y = 2x \\ 2x - 2y = 2y \end{cases} \Rightarrow \begin{cases} -x + 2y = 0 \\ 2x - 4y = 0 \end{cases}$$

$$x = 2y \Rightarrow \mathbf{s}_1 = \begin{bmatrix} 2t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} t \quad (t \neq 0)$$

•  $\lambda_2 = -3$

$$(\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} - \begin{bmatrix} -3 & 0 \\ 0 & -3 \end{bmatrix}) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} 4x + 2y = 0 \\ 2x + y = 0 \end{cases}$$

$$y = -2x \Rightarrow \mathbf{s}_2 = \begin{bmatrix} t \\ -2t \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} t \quad (t \neq 0)$$

- A sajátértékek és sajátvektorok függetlenek attól, hogy milyen bázisban felírt mátrixból számítjuk ki őket (invariánsak; a definícióból nyilvánvaló)
- Az  $n \times n$ -es valós szimmetrikus mátrixoknak minden van  $n$  db valós **sajátértéke**
- A szimmetrikus mátrixok különböző sajátértékeihez tartozó bármely sajátvektorai merőlegesek egymásra (vagyis a hozzájuk tartozó saját-alterek is merőlegesek)
  - Az előző példában:

$$s_1 \cdot s_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \end{bmatrix} = 0$$

## Szinguláris érték, szinguláris vektorok

- szingulársi érték jele:  $\sigma_i$
- szinguláris értékek a sajátértékek gyökei
  - $\sigma_i = \sqrt{\lambda_i}$

**DEF:**

- A pozitív  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  számok az  $r$ -rangú valós  $A$  mátrix szinguláris értékei:
  - ha a **sortérnek** van olyan  $(v_1; v_2; \dots; v_r)$  **ortonormált bázisa**
  - és az **oszloptérnek** van olyan  $(u_1; u_2; \dots; u_r)$  **ortonormált bázisa**, hogy:

$$Av_i = \sigma_i u_i; i = 1; \dots; r$$

- A  $v_i$  vektorokat jobb, az  $u_i$  vektorokat bal szinguláris vektoroknak nevezzük

### Mátrix rangja

- A mátrix Gauss-elimináció után kapott lépcsőfokok azaz a nem nulla nemnulla együtthatókkal rendelkező sorvektorok a mátrix rangja
- Példa:

$$\bullet A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 4 \\ 0 & 10 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 4 \\ 0 & 0 & -6 \end{pmatrix} \Rightarrow \text{rang}(A) = 3$$

$$\bullet B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 4 \\ 0 & 3 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 4 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \text{rang}(B) = 2$$

$$\bullet C = \begin{pmatrix} 2 & 3 \\ 0 & 1 \\ 4 & -1 \end{pmatrix} \sim \begin{pmatrix} 2 & 3 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow \text{rang}(C) = 2$$

Ortonormált bázis:

- Legyen  $V$  vektortér, amelyen definiálva van egy skaláris szorzat (azaz egy  $V \times V \rightarrow R$  szimmetrikus, bilineáris, pozitív definit függvény).
- Az  $e_1 \dots e_n$  vektorrendszert  $V$  **ortonormált bázisának nevezzük**, ha **minimális generátorrendszer**  $V$ -ben, minden vektor aegység hosszúságú és bármely két vektorra egymásra merőleges.

### Sortér, oszloptér

- Egy mátrix...
  - **oszlopvektorai** által kifeszített alteret **oszloptérnek**
  - **sorvektorai** által kifeszített alteret **sortérnek** nevezzük
- Példa: Az  $m \times n$ -es valós  $A$  mátrix **sortere**  $R^n$  vektorok által kifeszített altér, **oszloptere** pedig  $R^m$  vektorok alkotta altér

## Jelölés

$$\Sigma_1 = \text{diag}(\sigma_1; \sigma_2; \dots; \sigma_r) = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{bmatrix} \quad \mathbf{U}_1 = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_r] \quad \mathbf{V}_1 = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_r]$$

Mivel  $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ , ezért  $\mathbf{AV}_1 = \mathbf{U}_1 \Sigma_1$ , azaz

$$\mathbf{A} [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_r] = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_r] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{bmatrix}.$$

## Kiegészítés bázissá

Egészítük ki  $V_1$ -et és  $U_1$ -et a terek ortonormált bázisává:

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = [v_1 v_2 \dots v_r | v_{r+1} \dots v_n]$$

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = [u_1 u_2 \dots u_r | u_{r+1} \dots u_m]$$

## A teljes tér bázisával

Mivel  $\langle \mathbf{V}_2 \rangle = \text{Ker}(\mathbf{A})$ , ezért  $\mathbf{AV}_2 = \mathbf{0}$ . Mátrix alakban:

$$\mathbf{AV} = [\mathbf{Av}_1 \ \mathbf{Av}_2 \ \dots \ \mathbf{Av}_r \mid \mathbf{Av}_{r+1} \ \dots \ \mathbf{Av}_n] = [\sigma_1 \mathbf{u}_1 \ \sigma_2 \mathbf{u}_2 \ \dots \ \sigma_r \mathbf{u}_r \mid \mathbf{0} \ \dots \ \mathbf{0}]$$

Azaz

$$\mathbf{AV} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_r \mid \mathbf{u}_{r+1} \ \dots \ \mathbf{u}_m] \left[ \begin{array}{cccc|ccc} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{array} \right] = \mathbf{U}\Sigma.$$

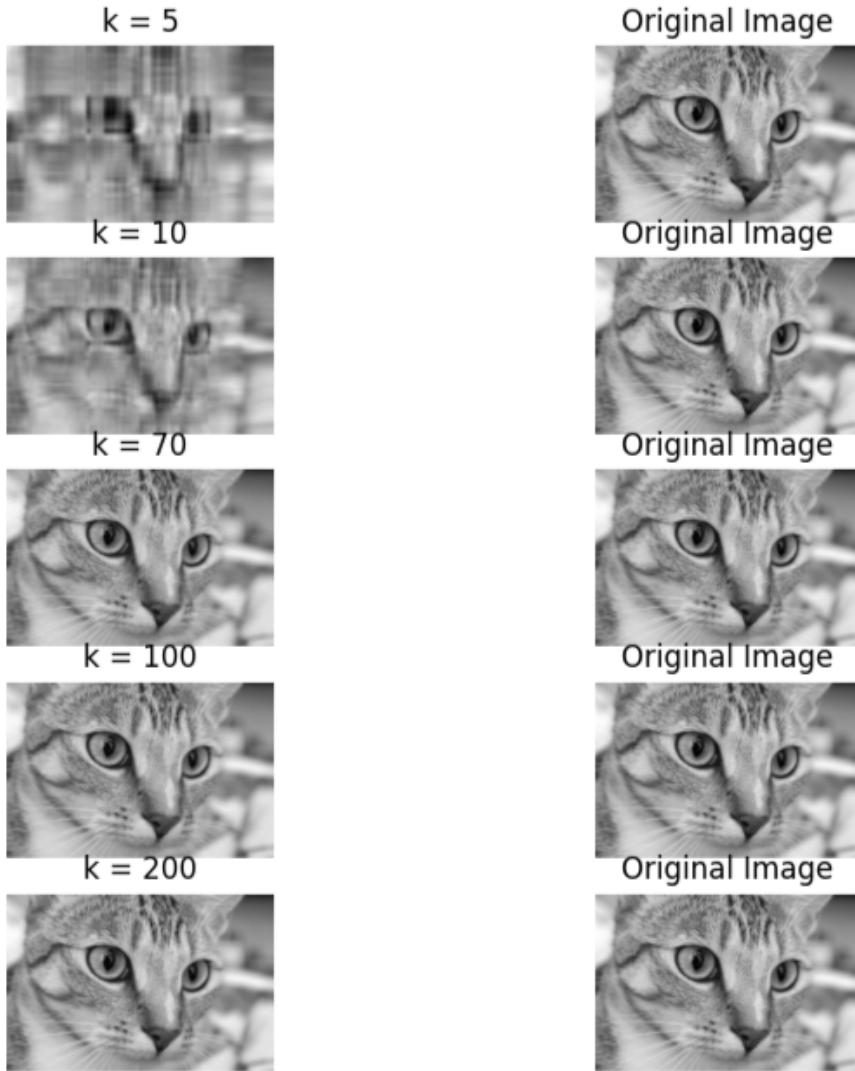
Méret szerint  $\mathbf{A}^{m \times n} \mathbf{V}^{n \times n} = \mathbf{U}^{m \times m} \Sigma^{m \times n}$ , vagy blokkmátrix alakban:

$$\mathbf{A} [\mathbf{V}_1 \mid \mathbf{V}_2] = [\mathbf{U}_1 \mid \mathbf{U}_2] \left[ \begin{array}{c|c} \Sigma_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right].$$

## Szinguláris érték felbontás (SVD)

- Olyan  $V$ -ket keresünk, hogy  $\sigma_j * u_i$  számszoros ( $i = \{1, 2 \dots n\}$ )
- Használatai:
  - Forgatás, tükrözés, nagyítás, kicsinyítés
  - Image Compression - kicsinyítés:

- Geeks4geeks Application of SVD - python implementációval együtt  
(<https://www.geeksforgeeks.org/singular-value-decomposition-svd/>)
  - „In this code, we will try to calculate the **Singular value decomposition** using **Numpy and Scipy**. We will be **calculating SVD**, and also performing **pseudo-inverse**. In the end, we can apply SVD for compressing the image“



- Adatbányászatban is használt - **Dimension Reduction**:
  - SVD magyarázó moodle-ből ([https://elearning.unibuda.hu/main/pluginfile.php/1132265/mod\\_resource/content/1/dim%20reduction%20with%20SVD.pdf](https://elearning.unibuda.hu/main/pluginfile.php/1132265/mod_resource/content/1/dim%20reduction%20with%20SVD.pdf))
  - Why should we reduce dimensions:
    - Discover hidden correlations/topics
    - Words that occur commonly together
    - Remove redundant and noisy features -> not all words are useful
    - Interpretation and visualization
    - Easier storage and processing of the data
- Az  $A$  mátrix szinguláris érték (szerinti) felbontása (SVD, singular value decomposition)

$$A = U\Sigma V^T$$

- $A$ : bemeneti mátrix  $m \times n$  pl (m dokumentumok n kifejezések)

- $U$ : Bal szinguláris vektorok =>  $m \times r$  matrix pl (m dokumentumok, r concepciók)
- $\Sigma$ : szinguláris értékek =>  $r \times r$  mátrix errőssége egyes koncepcióknak ahol az r a mátrix rangja
- $V$ : jobb szinguláris vektorok =>  $n \times r$  pl (n kifejezések, r koncepciók)
- **redukált szinguláris érték** (szerinti) felbontása pedig

$$A = U_1 \Sigma_1 V^T$$

$$\begin{bmatrix} 0 & -2 & 2 \\ -2 & 3 & -2 \\ 4 & -2 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix}$$

- 1. Meghatározzuk  $A^T A$  sajátértékeit ( $\lambda_i$ ) és sajátvektorait ( $s_i$ ).
- 2. a **sajátértékek gyökei a szinguláris értékek**:  $\sigma_i = \sqrt{\lambda_i}$
- 3. a **jobb szinguláris vektorok a sajátvektorok**:  $v_i = \frac{s_i}{|s_i|}$
- 4. A **bal szinguláris vektorok**:
  - 4.1 mivel  $Av_i = \sigma_i u_i$ , ezért  $u_i = \frac{1}{\sigma_i} Av_i$
  - 4.2 Másik lehetőség: a bal szinguláris vektorok **megegyeznek** az  $AA^T$  **sa játvektoraival**.
- 5. A szinguláris vektorokból felírható  $V_1$  és  $U_1$ , a **redukált szinguláris** felbontás.
- 6. A  $V$ -hez és az  $U$ -hoz ki kell egészíteni  $V_1$ -et és  $U_1$ -et ortonormált bázissá.
- Példa sajátérték és sajátvektor számításra

6

$$\begin{aligned}\varphi(s) &= \lambda s, \\ As &= \lambda s, \\ As - \lambda s &= 0, \\ (A - \lambda E)s &= 0, \\ \det(A - \lambda E) &= 0.\end{aligned}$$

$$A = \begin{bmatrix} 2 & -5 \\ -3 & 5 \end{bmatrix}$$

$$\lambda E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A - \lambda E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & -5 \\ -3 & 5 \end{bmatrix} \sim$$

- $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$

$$\det \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix} = \{(2-\lambda) \cdot (5-\lambda)\} - \{-3 \cdot (-3)\} =$$

$$10 + \lambda^2 - 5\lambda - \lambda^2 =$$

$$-10 + \lambda^2 - 5\lambda - 3 = \underline{\underline{\lambda^2 - 5\lambda + 1}}$$

$$\lambda^2 - 5\lambda + 1 = 0 \quad \left. \begin{array}{l} \lambda_1 = 6,85 \\ \lambda_2 = 0,14 \end{array} \right\} \leftarrow \text{magas értékek}$$

↓

$$\boxed{\begin{array}{l} \sigma_1 = \sqrt{6,85} \\ \sigma_2 = \sqrt{0,14} \end{array}}$$

- Levezetett feladatmegoldás SVD:

- BME-s SVD megoldás példafeladatokkal, magyarázatokkal  
([http://sandbox.hlt.bme.hu/~gaebor/ea\\_anyag/FelsobbMat/SVD.pdf](http://sandbox.hlt.bme.hu/~gaebor/ea_anyag/FelsobbMat/SVD.pdf))

- Példa feladat levezetéssel:

**0.2.1. Feladat.** Számítsuk ki az

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$$

mátrix SVD felbontását!

*Megoldás 1.* Számítsuk ki  $B = AA^\top$  sajátértékeit és sajátvektorait!

$$B = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}$$

Ennek sajátértékei:

$$\det \begin{pmatrix} 11 - \lambda & 1 \\ 1 & 11 - \lambda \end{pmatrix} = (11 - \lambda)^2 - 1 = \\ 120 - 22\lambda + \lambda^2$$

Ebből a sajátértékek:  $\lambda_1 = 10, \lambda_2 = 12$ . Vagyis a szinguláris értékek  $\sqrt{10}$  és  $\sqrt{12}$ .  $\lambda_1 = 10$ -hez sajátvektor (első jobb-szinguláris vektor):

$$\left[ \begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array} \right]$$

ennek egy nem-nulla megoldása a  $(-1, 1)$  vektor.

A 12 sajátértékhez tartozó sajátvektor:

$$\left[ \begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array} \right]$$

Vagyis a második jobb-szinguláris vektor az  $(1, 1)$ .

$$U' = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

A bal-szinguláris vektorok:

$$V' = A^\top U' = \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -4 & 2 \\ 2 & 4 \\ 0 & 2 \end{pmatrix}$$

$U$  és  $V$  úgy kapható meg, hogy  $U'$  és  $V'$  oszlopaikat normalizáljuk:

$$U = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, V = \begin{pmatrix} \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{5}} & \sqrt{\frac{2}{3}} \\ 0 & \frac{1}{\sqrt{6}} \end{pmatrix}, \Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{12} \end{pmatrix}$$

*Megoldás 2.* Számítsuk ki  $B = A^\top A$  sajátértékeit és sajátvektorait!

$$B = \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

Ennek sajátértékei:

$$\det \begin{pmatrix} 10 - \lambda & 0 & 2 \\ 0 & 10 - \lambda & 4 \\ 2 & 4 & 2 - \lambda \end{pmatrix} = (10 - \lambda)((10 - \lambda)(2 - \lambda) - 16) - 4(10 - \lambda) = \\ -\lambda \cdot (120 - 22\lambda + \lambda^2)$$

Ebből a sajátértékek:  $\lambda_1 = 0$ ,  $\lambda_2 = 10$ ,  $\lambda_3 = 12$ . Vagyis a szinguláris értékek  $0$ ,  $\sqrt{10}$  és  $\sqrt{12}$ . A nullához nem is keresünk sajátvektort, mert a végeredményben úgyis nulla szorzóval szerepelne.  $\lambda_2 = 10$ -hez sajátvektor (első jobb-szinguláris vektor):

$$\left[ \begin{array}{ccc|c} 0 & 0 & 2 & 0 \\ 0 & 0 & 4 & 0 \\ 2 & 4 & -8 & 0 \end{array} \right]$$

ennek egy nem-nulla megoldása a  $(-2, 1, 0)$  vektor.

A 12 sajátértékhez tartozó sajátvektor:

$$\left[ \begin{array}{ccc|c} -2 & 0 & 2 & 0 \\ 0 & -2 & 4 & 0 \\ 2 & 4 & -10 & 0 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} -2 & 0 & 2 & 0 \\ 0 & -2 & 4 & 0 \\ 0 & 4 & -8 & 0 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} -2 & 0 & 2 & 0 \\ 0 & -2 & 4 & 0 \\ 0 & 1 & -2 & 0 \end{array} \right] \rightarrow$$

$$\left[ \begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -2 \end{array} \right]$$

Vagyis a második jobb-szinguláris vektor a  $(1, 2, 1)$ .

$$V' = \begin{pmatrix} -2 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Ezt a mátrixot kiegészíthetnénk egy harmadik, ez előző kettőre ortogonális vektorral és bevehetnénk a 0 saját-értéket. Ekkor kapnánk teljes SVD-t, de ezt most nem tesszük meg.

A bal-szinguláris vektorok:

$$U' = AV' = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -5 & 6 \\ 5 & 6 \end{pmatrix}$$

$U$  és  $V$  úgy kapható meg, hogy  $U'$  és  $V'$  oszlopaikat normalizáljuk:

$$U = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{5}} & 0 \end{pmatrix}, V = \begin{pmatrix} \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{5}} & \sqrt{\frac{2}{3}} \\ 0 & \frac{1}{\sqrt{6}} \end{pmatrix}, \Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{12} \end{pmatrix}$$

Ugyan az jött ki, mint az első megoldásban, bár ez nem szükségszerű, mivel a felbontás nem egyértelmű.  
Szinguláris felbontás alakban:

$$A = \sqrt{10} \cdot \underbrace{\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}}_{\text{diád szorzat}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} + \sqrt{12} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}}_{\text{diád szorzat}} \begin{pmatrix} \frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

## SVD HATÁSA

- Legyen  $A$  egy  $r$ -rangú,  $m \times n$ -es valós mátrix.
- Az  $x \rightarrow Ax$  leképezés az  $R_n$  tér  $e^T e = 1$  egyenletet kielégítő egységgömb felületén lévo pontjait az  $R^m$  tér egy  $r$ -dimenziós alterének...
  - egy **ellipszoidjának felületére** képezi, ha  $r = n$
  - egy **ellipszoidja által határolt** tartományára képezi, ha  $r < n$

## OPTIMALIZÁLÁS

### Mátrixok definiáció

Mátrix definiáció fogalma & megoldási módszerek leírása (<https://www.unimiskolc.hu/~matente/oktatasi%20tananyagok/M%C3%81TRIX%20DEFINITS%C3%89G%C3%89NEK%20FOGALMA%20%C3%89S%20TESZTEK%20A%20DEFINITS%C3%89G%20ELD%C3%96NT%C3%89S%C3%89RE.pdf>)

- $x^T$  az  $x$  mátrix transzponáltját jelenti

- Az  $A \in R^{n \times n}$  mátrix **pozitív definit**, ha

$$x^T Ax > 0,$$

$$[x^T Ax < 0],$$

$$(x \in R^n, \forall x \neq 0)$$

- Az  $A \in R^{n \times n}$  mátrix **pozitív szemidefinit**, ha

$$x^T Ax \geq 0,$$

$$[x^T Ax \leq 0],$$

$$(\forall x \in R^n)$$

- Az  $A \in R^{n \times n}$  mátrix **indefinit**, ha egyik fenti kategóriába sem tartozik.
- Megjegyzés.: Az A mátrix pontosan akkor **negatív [szemi]definit**, ha  $-A$  pozitív [szemi]definit.
- Az  $[a_{ij}]_{i,j=1}^n \in R^{n \times n}$  mátrix  **$k$ -adik főminormátrixa** az  

$$\mathbf{A}_k = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{bmatrix} \quad (k = 1; \dots; n)$$
- Tétel: A szimmetrikus  $A \in R^{n \times n}$  mátrix pontosan akkor **pozitív [negatív] definit**, ha  
 $\det \mathbf{A}_i > 0, \quad i \in \{1; \dots; n\};$   
 $[\operatorname{sgn}(\det \mathbf{A}_i) = (-1)^i, \quad i \in \{1; \dots; n\}]$

- **Tételek:**

- Az  $A \in R^{n \times n}$  szimmetrikus mátrix **pontosan akkor pozitív [negatív] definit**, ha  $A$  minden sajátertéke pozitív[negatív] (valós) szám.
- Az  $A \in R^{n \times n}$  szimmetrikus mátrix pontosan akkor pozitív [negatív] szemidefinit, ha A minden sajátertéke nem-negatív [nem-pozitív] (valós) szám.
  - nem-negatív: 0, vagy pozitív(sajátertékek  $\geq 0$ )
  - nem-pozitív: 0, vagy negatív(sajáterték  $\leq 0$ )
- Az  $A \in R^{n \times n}$  szimmetrikus mátrix **pontosan akkor pozitív definit**, ha az  $A = LU$  felbontásában
  - ahol  $L$  egység alsóháromszög-mátrix és  $U$  elsőháromszög-mátrix
  - az  $U$  mátrix minden diagonális eleme pozitív
  - **diagonális elem:** a főátlón található elemek
  - A háromszögmátrix olyan négyzetes mátrix, melynek a főátlója alatti összes elem vagy a főátlója feletti összes elem zéró.

- Megoldási módok (lásd fentebb a linket részletes levezetésért):

- Gauss-módszeres megoldás:
  - Gauss-módzserrel előállítjuk az alsó háromszög mátrixunkat ezzel együtt kijön az  $U$
  - **Pozitív definit:** a felső  $U$  háromszögmátrixba bele kell venni a főátló elemit és csak ezeknek az elemeknek kell pozitívnak lenniük. Ezek alapján a lenti mátrix eleget tesz a definícióknak

$$U = \begin{bmatrix} 4 & -2 & 2 \\ 0 & 4.5 & -3 \\ 0 & 0 & 1.6 \end{bmatrix}$$

- **Negatív definit:** Ha  $A$  mátrix  $-1$ -szeresét vesszük, és a főátló elemei mind pozitívak, akkor az  $A$  mátrix

- Főminor megoldási mód:

Az  $\mathbf{A}_k$  ( $k = 1, 2, 3$ ) főminormátrixok és azok determinánsai, a főminorok az alábbiak:

$$\mathbf{A}_1 = \begin{bmatrix} 2 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -6 \\ -2 & -6 & 10 \end{bmatrix}$$

$$\det(\mathbf{A}_1) = 2, \quad \det(\mathbf{A}_2) = \cancel{X}^2, \quad \det(\mathbf{A}_3) = \det(\mathbf{A}) = 8.$$

## Többváltozós deriváltak

- Több változós deriváltnál azokat a változókat amikkel nem deriválunk konstansnak vesszük, pl ha xés y változónk van és x-szerint kell deriválni akkor y-t konstansnak kell kezelni

- ha csak egy változó alapján deriválunk az elsőrendű parciális deriválásnak hivjuk

$$x \text{ szerint: } \frac{\partial}{\partial x} (x^2 + 2y) = 2x$$

$$y \text{ szerint: } \frac{\partial}{\partial y} (x^2 + 2y) = 2$$

- Kétszeresen deriválunk azt másodrendű parciális deriválásnak nevezzük
  - PL  $f'_{xy}$  vagyis először x aztán y szerint deriválunk akkor
    - $\frac{\partial}{\partial x} (x^2 + 3yx) = 2x + 3y$
    - $\frac{\partial}{\partial y} (2x + 3y) = 3$
    - 1. először x szerint: y-t számnak vagyis konstansnak vesszük és lederiválunk
    - 2. y szerint: az előző deriválás eredményét lederiváljuk úgy hogy az x-eket konstansnak vesszük
- $\nabla$ : ez a szar a 'Nebla', jelentése ismeretlen :D, a vadonban, matematikai és fizikai papírokban viszonylag gyakran előfordul a gradiens vektorok téma körében
- Az  $f : R^n \rightarrow R$ , n-változós valós függvény **gradiense** (graidentvektora)

$$\nabla f(x) = \left[ \frac{\partial f(x)}{\partial x_1}; \dots; \frac{\partial f(x)}{\partial x_n} \right]^T = \nabla_x f(x)$$

- Az  $f : R^n \rightarrow R$ , n-változós valós függvény Hesse-mátrixa

$$\nabla^2 f(\mathbf{x}) = \left[ \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right]_{i,j=1}^n = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix} = \nabla_{\mathbf{x}\mathbf{x}}^2 f(\mathbf{x})$$

- **Hesse mátrix**: egy többváltozós valós függvény másodrendű parciális deriváltjaiból alkotott négyzetes mátrixát nevezzük. Lényegében ez a szar itt (ugyan az mint a fenti ábra, csak anal jegyzetből van, nem dimat):

$$\begin{bmatrix} f'_{xx}(P_0) & f'_{xy}(P_0) \\ f'_{yx}(P_0) & f'_{yy}(P_0) \end{bmatrix}$$

### Gradiensmátrix:

- 1-m változót szerint derivál le és egymás mellé rakja az eredményt

$$\begin{aligned} F : R^n &\rightarrow R^m \\ F(x) &= [F_1(x); \dots; F_m(x)], \\ F_i &: R^n \rightarrow R \end{aligned}$$

- vektor-vektor függvény gradiense (gradiensmátrixa) a

$$\nabla F(x) = [\nabla F_1(x); \dots; \nabla F_m(x)] \in R^{n \times m}$$

- Az  $f(x)$  mennyisége  $O(g(x)) = \text{ordó } gx$  nagyságrendű
  - ha  $\|f(x)\| \leq K \|g(x)\|$  egyenlőség igaz valamely  $K > 0$  konstans esetén

## Taylor-sorok:

- Taylor-sor összegfüggvénye (anal2 definíció kicsit érthetőbb mint a lenti borzalom :D) ahol az  $f^{(k)}$  a f-adik deriváltat jelöli

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(a) \frac{(x-a)^k}{k!}$$

- A **Taylor-sor magyarúl**: hatványsorokkal megközelíteni egy bonyolult függvényt
- Taylor sor formális definíciója

Az  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  függvény  $\mathbf{x}_0$  körüli Taylor-sora

$$f(\mathbf{x}_0 + \mathbf{p}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \nabla^2 f(\mathbf{x}_0) \mathbf{p} + \dots + \frac{1}{r!} D^r f(\mathbf{x}_0) + \dots$$

ahol  $\mathbf{x}_0, \mathbf{p} \in \mathbb{R}^n$ , és

$$D^s f(\mathbf{x}_0) = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_s=1}^n \left( p_{i_1} p_{i_2} \dots p_{i_s} \frac{\partial^s f(\mathbf{x}_0)}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_s}} \right).$$

- Taylor polinom szélsőértékhely kapcsolata
  - A Taylor-polinomokat használhatjuk szélsőértékek meghatározására is. ha egy függvénynek van lokális minimuma vagy maximuma egy adott pontban, akkor az első deriváltjának értéke 0 lesz abban a pontban. A második derivált értékének előjele megmutatja, hogy minimumról vagy maximumról van-e szó. Ha a második derivált pozitív, akkor minimumról van szó, ha negatív, akkor maximumról. Ha 0, akkor további vizsgálat szükséges. (feladat kérdezte :( )

Két speciális esetet különítünk el:

1.eset :

- Az  $f: R^n \rightarrow R$  függvény  $x_0 \in R^n$  pontbeli lineáris közelítése az f elsorendű Taylor-polinomja:

$$\begin{aligned} f(x_0 + p) &\approx f(x_0) + \nabla f(x_0)^T p, \\ p &\in R^n \end{aligned}$$

- A közelítés hibája  $O(\|p\| 2)$ , ha f **kétszer folytonosan** differenciálhat:

$$y = y(p) = f(x_0) + \nabla f(x_0)^T p$$

2.eset.:

- Az  $f : R^n \rightarrow R$  függvény  $x_0 \in R^n$  pontbeli **kvadratikus közelítése** az f másodrendű Taylor-polinomja

$$f(x_0 + p) \approx f(x_0) + \nabla f(x_0)^T p + \frac{1}{2} p^T \nabla^2 f(x_0) p, \quad (p \in R^n)$$

- A közelítés hibája  $O(\|p\|^3)$ , ha f háromszor folytonosan differenciálható.
- Az

$$y = f(x_0) + \nabla f(x_0)^T p + \frac{1}{2} p^T \nabla^2 f(x_0) p$$

másodfokú függvény az f függvény grafikonjának érintő paraboloidja az  $\frac{x_0}{f(x_0)} \in R^{n+1}$  pontban

## Függvények szélsőértéke

### Abszolút szélsőérték hely

- Az  $f : R^n \rightarrow R$  ( $n \geq 1$ ) függvény abszolút (globális) minimumhelye [maximumhelye] az  $x^* \in D(f)$  pont, ha

$$f(x^*) \leq f(x) \quad [f(x^*) \geq f(x)] \quad \forall x \in D(f).$$

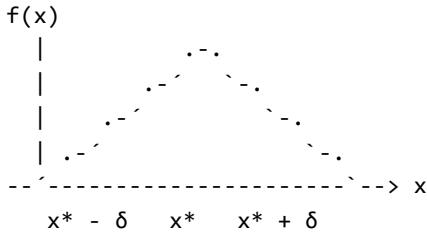
- **Magyarul**

- $x^*$ : abszolút minimum vagy maximumhely jelölése
- $D(x)$ : Értelmezési tartomány
- $x^*$  minimumhely ha:  $f(x^*) \leq f(x)$
- $x^*$  maximumhely ha:  $f(x^*) \geq f(x)$

### Gömbi környezet

- Az  $S(x^*; \delta) = \{x \in R^n \mid \|x - x^*\| < \delta\} \subset R^n$  halmaz az  $x^* \in R^n$  pont  $\delta$  sugarú ( $\delta > 0$ ) nyílt (gömbi) környezete
- remélhetőleg erről a környezetről van szó
- Az  $S(x^*; \delta)$  az  $x^*$  pont  $\delta$  sugarú környezetét jelenti. Ez azt jelenti, hogy az összes olyan pontot tartalmazza, amelyek távolsága az  $x^*$  ponttól kisebb, mint  $\delta$ . Más szóval, ez az összes olyan pont halmaza, amelyek a következő feltételnek megfelelnek:  $\|x - x^*\| < \delta$ , ahol  $\|\cdot\|$  az euklideszi távolságot jelenti.

### Lokális szélsőérték hely



- Az  $f : R^n \rightarrow R$  ( $n \geq 1$ ) függvény lokális **minimum [maximum]helye** az  $x^* \in D(f)$  pont, ha létezik  $\delta > 0$ , hogy

$$\begin{aligned} f(x^*) &\leq f(x) \\ [f(x^*) &\geq f(x)] \\ \forall x \in D(f) \cap S(x^*; \delta). \end{aligned}$$

#### Erős/szigorú szélsőértékhely

- Az  $x^* \in D(f)$  **minimum [maximum]hely** erős (szigorú), ha valamely  $\delta > 0$  esetén

$$\begin{aligned} f(x^*) &< f(x), \\ [f(x^*) &> f(x)] \\ \forall x \in D(f) \cap S(x^*; \delta), x \neq x^* \end{aligned}$$

#### Gyenge szélsőértékhely

- Ha a szélsoérthely nem erős (szigorú), akkor **gyenge**.

#### Szélsőérték

- A szélsoérthelyen (minimumhelyen/maximumhelyen) felvett függvényérték a szélsőérték (minimumérték/maximumérték).

#### Egyváltozós

##### Elsorendű szükséges feltétel (szélsoérthete van)

- Ha  $f : R \rightarrow R$  függvénynek  $x^*$  pontban szélsoérthete van, és  $f$  folytonosan differenciálható  $x^*$  pontban, akkor  $f'(x^*) = 0$ .

##### Másodrendű szükséges feltétel (minimuma [maximuma] van)

- Ha  $f : R \rightarrow R$  függvénynek  $x^*$  pontban minimuma [maximuma] van, és  $f$  kétszer folytonosan differenciálható  $x^*$  pontban, akkor

$$\begin{aligned} f'(x^*) &= 0 \\ \text{és} \\ f''(x^*) &\geq 0 \text{ vagy } [f''(x^*) \leq 0] \end{aligned}$$

## Másodrendű elégséges feltétel

- Ha  $f : R \rightarrow R$  függvény kétszer folytonosan differenciálható  $x^*$  pontban,  $f'(x^*) = 0$  és  $f''(x^*) > 0$  [ $f'(x^*) < 0$ ], akkor  $x^*$  az  $f$  függvény minimumhelye [maximumhelye].

## Többváltozós

### Stacionárius pont

- Először meghatározzuk a szélsőértékeket és ha ezeket behelyettesítve  $x$  és  $y$  alapján derivált függvényekbe és 0-t kapunk az a szélsőérték egy stacionárius pont is
- Lépések:
  - $x$  és  $y$  alapján lederiváljuk a fügvényt
  - Szélsőértékek meghatározása ( $P_0$  pontok): A deriválás eredményéül kapott fügvényeket egyenlővé tesszük 0-val vagyis megnézzük hogy milyen értéket kell hogy  $x$  és  $y$  felvegyen hogy behelyettesítés után 0-t kapunk  $\Rightarrow f(x,y)=0$  akkor  $P_0(x,y)$
  - Ha egy szélsőérték  $x,y$  által derivált fügvénynél is jelen van az a szélsőérték egy stacionárius pont

7)  $f(x) = 8x_1 - x_1^2 x_2 + x_2^2$

$$\frac{\partial}{\partial x_1} = 8 - 2x_1 x_2 \rightsquigarrow 8 - 2x_1 y = 0 \Rightarrow$$

$$\frac{\partial}{\partial x_2} = -x_1^2 + 2x_2 \rightsquigarrow -x_1^2 + 2y = 0$$

$$\begin{cases} 8 - 2x_1 y = 0 \\ -x_1^2 + 2y = 0 \end{cases} \quad \begin{matrix} 2x_1 y = 8 \\ x_1^2 = 2y \end{matrix} \quad \underbrace{P_0(2,2)}_{P_0(1,1)}, P_0(1,1), P_0(-1,1)$$

$$-x_1^2 + 2y = 0 \quad P_0(0,0), P_0(2,2) \text{ STAC}$$

## Elsőrendű szükséges feltétel

- Ha  $f : R^n \rightarrow R$  függvénynek  $x^*$  pontban szélsőértéke van, és  $f$  folytonosan differenciálható  $x^*$  pontban, akkor:  $\nabla f(x^*) = 0$

## Terminológia

- A  $\nabla f(x^*) = 0$  elsőrendű szükséges feltétel elnevezése **stacionárius egyenlet**.
-

## Másodrendű szükséges feltétel

- Ha  $f : R^n \rightarrow R$  függvénynek  $x^*$  pontban minimuma [maximuma] van, és  $f$  kétszer folytonosan differenciálható  $x^*$  pontban, akkor  $\nabla f(x^*) = 0$ , és a  $\nabla^2 f(x^*)$  Hesse-mátrix pozitív [negatív] szemidefinit.

## Másodrendű elégsges feltétel

- Ha  $f : R^n \rightarrow R$  kétszer differenciálható  $x^*$  pontban,  $\nabla f(x^*) = 0$ , és a  $\nabla^2 f(x^*)$  Hesse-mátrix pozitív [negatív]definit, akkor  $x^*$  az  $f$  függvény minimumhelye [maximumhelye].

## Megjegyzések

- A Hesse-mátrix definiitsege a függvény adott pontbeli (lokális) görbületségével van kapcsolatban. A Taylor-sor alapján:
  - Ha  $\nabla^2 f(x^*)$  pozitív szemidefinit, akkor az  $f$  függvény  $x^*$  pontban lokálisan konvex (ha pozitív definit, akkor szigorúan konvex), mert grafikonja az adott pontban az érintosíkhoz képest minden irányban felfelé görbü).
  - Ha  $\nabla^2 f(x^*)$  negatív szemidefinit, akkor az  $f$  függvény  $x^*$  pontban lokálisan konkáv (ha negatív definit, akkor szigorúan konkáv), mert grafikonja az adott pontban az érintosíkhoz képest minden irányban lefelé görbü)
- A szükséges feltétel szemléletesen azt állítja, hogy minimumhelyen [maximumhelyen] a függvény grafikonjának érintosíkja vízszintes, és a függvény (lokálisan) konvex [konkáv].
- Az elégsges feltétel szemléletesen azt állítja, hogy ha egy pontban a függvény grafikonjának érintosíkja vízszintes, és ott a függvény (lokálisan) szigorúan konvex [szigorúan konkáv], akkor a pont erős (szigorú) minimumhely [maximumhely].

## Megjegyzések 2

- A  $\nabla f(x) = 0$  ( $f : R^n \rightarrow R$ ) egyenlet megoldását általában az  $f$  **stacionárius** pontjának, ritkábban kritikus pontjának nevezik.
- Az  $f$  függvény  $x^*$  **stacionárius** pontja nem-elfajult (nem-degenerált), ha  $\det(\nabla^2 f(x^*)) \neq 0$ .
- Az  $f$  függvény  $x^*$  nem-elfajuló **stacionárius** pontjaiban a  $\nabla^2 f(x^*)$  Hesse-mátrixnak  $\ell$  pozitív és  $n - \ell$  negatív sajátértéke van ( $0 \leq \ell \leq n$ ), azaz nincs 0 sajátértéke.
  - Ha  $\ell = 0$ , akkor  $\nabla^2 f(x^*)$  negatív definit ( $x^*$  maximumhely).
  - Ha  $\ell = n$ , akkor  $\nabla^2 f(x^*)$  pozitív definit ( $x^*$  minimumhely).
  - Ha  $0 < \ell < n$ , akkor **Morse**-féle  $\ell$ -nyeregrol, vagy egyszerűen **nyeregpontról** beszélünk.
- Nyeregpontban nincs szélsoértek.
- Elfajult (degenerált) **stacionárius** pontban azonban lehet szélsoértek (lásd, pl.  $f(x_1; x_2) = x_1^4 + x_2^4$  a  $x^* = [0; 0]^T$  pontban), de ezek vizsgálata bonyolult.

# NUMERIKUS MINIMUMKERESŐ ELJÁRÁSOK

- Numerical Optimization Techniques - Princeton  
(<https://www.cs.princeton.edu/courses/archive/spring10/cos424/slides/6-opt.pdf>) Boncolgatott témák:
  - Line search
  - Coordinate-Gradient-Steepest Descent
  - Hessian matrix
  - Newton method
  - Conjugate Gradient algorithm
  - Stochastic Gradient Descent
- Prog1-ben tanult Logaritmikus kereséshez hasonló módszerek. Legalábbis felfogás szempontjából:
  - folyamatosan egyre jobban közelítjük a megoldást
    - vagy az intervallumok szűkítésével
    - vagy a fv tulajdonságait kihasználva egyéb matematikai megoldásokkal

## Egyváltozós függvények numerikus minimumkereső eljárásai

### Direkt és indirekt kereső eljárások

- **Direkt** kereso eljárások:
  - Egy  $[a_0; b_0]$  intervallumból indul, amely tartalmazza az  $x^*$  minimumpontot(befoglaló intervallum)
  - A kezdeti befoglaló intervallumot szűkíti az eljárás addig, amíg a megfelelo pontosságot el nem éri.
- **Indirekt** kereso eljárások:
  - Az  $f'(x) = 0$  (többváltozós esetben a  $\nabla f(x) = 0$ ) stacionárius egyenletet oldja meg numerikusan

## Egyváltozós függvények direkt eljárásai

### Unimodális függvény

- **unimodális:** Szimmetrikus, szabályos

- Az  $f : [a; b] \rightarrow R$  függvény unimodális, ha bármely  $x_1, x_2 \in [a; b], x_1 < x_2$  pont esetén

$$\begin{aligned} x_2 \leqslant x^* &\Rightarrow f(x_1) > f(x_2), \\ x^* \leqslant x_1 &\Rightarrow f(x_1) < f(x_2) \end{aligned}$$

- **Szemléletesen:** ha  $f$  **unimodális** és  $x^*$  a minimumpontja, akkor  $f$  az  $x^*$ -tól **balra szigorúan monoton fogy** (igen a tanár így írta :D), míg **tőle jobbra szigorúan monoton nő**

### Szigorúan konvex függvény

- Az  $f : [a; b] \rightarrow R$  függvény szigorúan konvex, ha bármely  $x, y \in [a; b], x \neq y$  esetén

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y), \quad \forall \lambda \in (0; 1).$$

- Észrevétel: A szigorúan konvex függvények unimodálisak.

## Megállapítás

- Vegyünk két közbülső pontot:  $a < c < d < b$ . Ha  $f$  unimodális, akkor
  - ha  $f(c) < f(d)$ , akkor  $x^* \in [a; d]$ ;
  - ha  $f(c) \geq f(d)$ , akkor  $x^* \in [c; b]$

## Szemléletesen

- A befoglaló intervallum bal vagy jobb végpontját aszerint lehet „beljebb húzni”, hogy a két közbülső pontban felvett értékkel biztosítani lehessen, hogy a függvény megtartja unimodális jellegét (az „U” alakját).
- Egyetlen közbülső pont felhasználásával nem lehet az intervallumot szűkíteni.

## Általános kereső eljárás

```

1  input [a, b], ε > 0
2  repeat
3      select c, d ∈ (a, b) such that c < d.
4      if f(c) < f(d)
5          b = d
6      else
7          a = c
8      end
9  end until b - a < ε

```

- Közbülső pontok megválasztása:
  - Az egyik belső pont az új intervallum végpontja lesz, a másik pedig belső pontja.
  - Az utóbbit már egyszer kiszámított függvényértéket jó lenne felhasználni a következő ciklusban, gyorsítva ezzel az eljárást.
  - Kellene, hogy az új belső pont az új intervallumban is „jó” helyen legyen, vagyis a régi-új belső pont ugyanolyan arányban ossza az új intervallumot, mint a régit.
  - **Megoldás: aranymetszés**

## Aranymetsző minimumkereső eljárás

- Az aranymetszés aránya:

$$\varphi = \frac{1 + \sqrt{5}}{2} \approx 1,618$$

- a két osztópont következő arányban osztja az intervallumot:

$$\begin{aligned}\tau_2 &= \frac{1}{\varphi} = \frac{\sqrt{5} - 1}{2} \approx 0,618 \\ \tau_1 &= 1 - \tau_2 = \frac{3 - \sqrt{5}}{2} \approx 0,382\end{aligned}$$

- Megjegyzés:  $\tau_1 < \tau_2$
- algoritmusa:  
Set  $\tau_1 = \frac{3-\sqrt{5}}{2} \approx 0.382$ ,  $\tau_2 = \frac{\sqrt{5}-1}{2} \approx 0.618$ .

```

1 Input [a, b], ε > 0
2 c = a + τ_1 (b - a), Fc = f(c)
3 d = a + τ_2 (b - a), Fd = f(d)
4 while b - a ≥ ε do
5     if F < F_d
6         b = d, d = c
7         c = a + τ1 (b - a)
8         F_d = F_c , F_c = f(c)
9     else
10        a = c, c = d
11        d = a + τ2 (b - a)
12        F_c = F_d , F_d = f(d)
13    end
14 end

```

## Konvergencia

- Világos, hogy a kiszámított intervallumsorozat kielégíti a

$$[a_0; b_0] \supset [a_1; b_1] \supset \dots \supset [a_k; b_k] \supset \dots$$

relációláncot úgy, hogy minden egyes  $[a_k; b_k]$  intervallum tartalmazza az  $x^*$  minimumpontot.

- Ha  $b_k - a_k < \varepsilon$ , akkor az  $[a_k; b_k]$  bármelyik pontja választható az  $x^*$  minimumpont legfeljebb  $\varepsilon > 0$  hibájú közelítésének.
- A **legjobb közelítés** azonban  $\tilde{x} = \frac{a_k+b_k}{2}$ ,
  - amelynek hibája  $|x - x^*| < \frac{\varepsilon}{2}$
- Az aranymetsző eljárás **konvergenciája lineáris**.
- Az aranymetszo eljárást alkalmazzák nem-unimodális függvényre is, de ebben az esetben a konvergencia nem garantálható

## Newton-módszer (Newton–Raphson-módszer; érintő módszer)

- Előnye: Gyors
- Hátránya: Ki kell számolni a második deriváltat és ennek nehéz a kiszámítása, ha nem ismerjük a képletét
- Analízis I. ismétlés. Az  $f(x) = 0$  egyenlet gyökközelítő sorozata:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

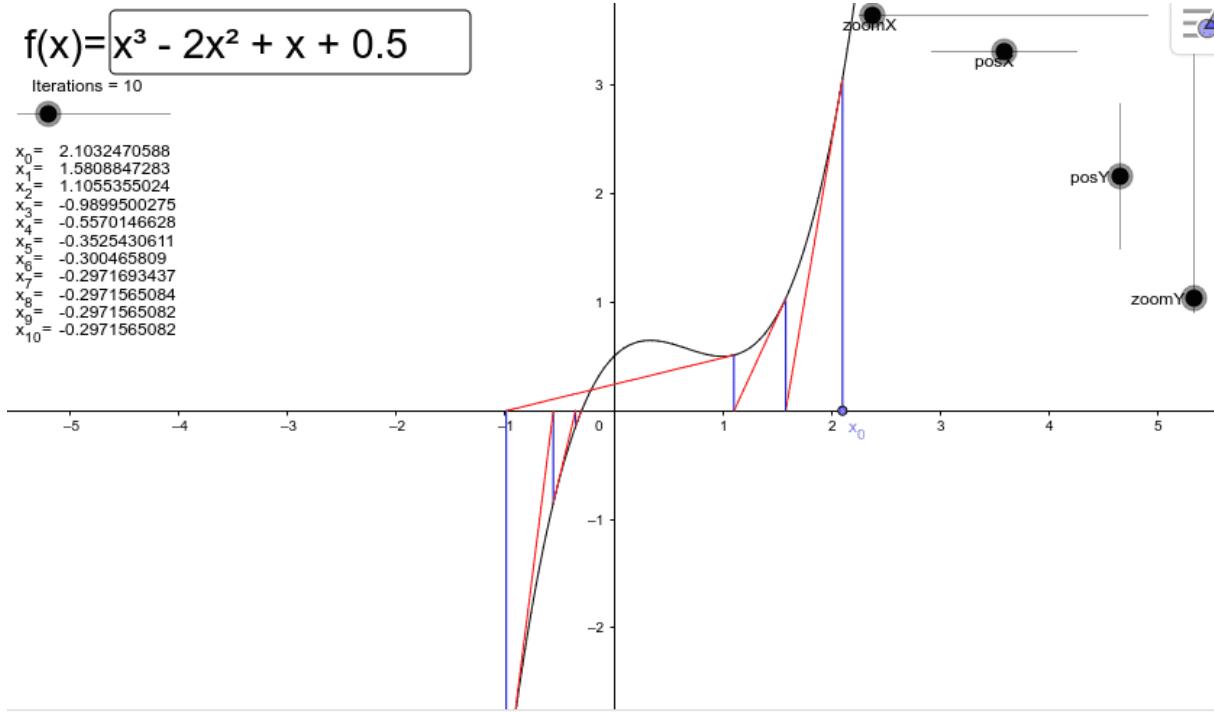
- Ha az  $f$  szélsőértékhelyét keressük, akkor az  $f'(x) = 0$  stacionárius egyenlet gyökközeliítő sorozata:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

- Megállási feltétel: ha a Newton-lépés elég kicsi, azaz

$$|x_{n+1} - x_n| = \frac{|f'(x_n)|}{|f''(x_n)|} < \varepsilon$$

- Newton-Raphson Method: GeoGebra (<https://www.geogebra.org/m/DGFGBJyU>)



## Többváltozós függvények numerikus minimumkereső eljárásai

- $\nabla f(x)$ : első derivált vektor formában és egy pont behelyettesítve
  - $f(x) = x_1^3 + 2x_2^2; (1; 1)$
  - $\frac{d}{dx_1} = 3x_1^2; \frac{d}{dx_2} = 4x_2$
  - $\nabla f(x) = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$
- Ha a többváltozós  $f$  szélsőértékhelyét keressük, a  $\nabla f(x) = 0$  **stacionárius** egyenlet gyökközeliítő sorozata:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

- Az inverzmátrix kiszámítása igen költséges, ezért inkább az

$$s_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad \text{Newton-lépést a}$$

$$\nabla^2 f(x_k) s_k = -\nabla f(x_k)$$

lineáris egyenletrendszerből számítjuk ki.

- A Hesse-mátrix szimmetrikus, szerencsés esetben ( $f$  szigorúan konvex) **pozitív definit**
- Az ilyen egyenletrendszer megoldására léteznek hatékony megoldások
- Newton algoritmus:

**repeat**

$$\nabla_{\mathbf{x}\mathbf{x}}^2 f(\mathbf{x}_k) \mathbf{s}_k = -\nabla f(\mathbf{x}_k)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$$

**end until**  $\|\mathbf{s}_k\| < \varepsilon$

## Kvázi-Newton-módszerek

- A kvázi-Newton-módszerek a Newton-módszerek iterációinkénti alacsony számítási költségű, mégis nagy hatékonyságú módosításai. Az alapötlet a  $\nabla^2 f(x_k)$  Hesse-mátrix helyettesítése egy szimmetrikus, pozitív definit  $H_k$  mátrixszal, amely jól közelíti a Hesse-mátrixot, és a számítási költsége alacsony. A következő algoritmus az egyik leghatékonyabb a numerikus eljárások között.

### BFGS (Broyden–Fletcher–Goldfarb–Shanno) algoritmus

for  $k = 0, 1, \dots$

$$\begin{aligned} \mathbf{s}_k &= -\mathbf{H}_k^{-1} \nabla f(\mathbf{x}_k) \\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{s}_k \\ \mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k) \\ \mathbf{H}_{k+1}^{-1} &= \left( \mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) \mathbf{H}_k^{-1} \left( \mathbf{I} - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \end{aligned}$$

end

- A  $H_0$  a  $\nabla^2 f(x^*)$  egy közelítése, de gyakran egyszerűen az  $n \times n$ -es egységmátrixot választják

## Vonalmenti minimumkereső eljárások

### Általános alak

- A vonalmenti keresés általános algoritmusa:
  - Az  $x_k$  iterációs pontban találunk keresési irányt:  $s_k$ .
  - Menjünk ebben az irányban valamennyit:  $x_{k+1} = x_k + \alpha_k s_k$ .
- Természetesen, nem mindegy, hogy milyen irányban és mennyit megyünk
  - Rögzítsük, hogy  $s_k$  legyen csökkenési irány:  $\nabla f(x_k)^T s_k < 0$  (tompaszöget zárnak be).
  - Ekkor  $\alpha_k \geq 0$ , amelynek neve learning rate

- Minimalizáljuk a  $g(\alpha) = f(x_k + \alpha s_k)$  egyváltozós függvényt az  $\alpha \geq 0$  tartományon:  $\alpha_k = \text{argmin} g(\alpha)$ . (Azaz haladjunk csökkenési irányban, egyenes mentén addig, amíg csökkenést tapasztalunk. Ott válasszunk másik irányt.)
- Vonalmenti minimumkereső

```

Input  $\mathbf{x}_0$ 
for  $k = 0, 1, 2, \dots$ 
    Select  $\mathbf{s}_k \in \mathbb{R}^n$ 
     $\alpha_k = \arg \min_{\alpha \geq 0} f(\mathbf{x}_k + \alpha \mathbf{s}_k)$ 
     $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{s}_k$ 
end

```

Keresési irány megválasztása

irány:  $\mathbf{s}_k = -\nabla f(\mathbf{x}_k)$ .

$\Leftarrow \mathbf{s}_k = -\mathbf{H}_k^{-1} \nabla f(\mathbf{x}_k)$ , ahol  $\mathbf{H}_k$  szimmetrikus pozitív definit mátrix.

módszer konkrétan:  $\mathbf{H}_k = \nabla^2 f(\mathbf{x}_k)$  (és  $\alpha_k = 1$ )

$\mathbf{s}_k$  keresési irányok a  $\pm \mathbf{e}_1, \pm \mathbf{e}_2, \dots, \pm \mathbf{e}_n \in \mathbb{R}^n$

$$\mathbf{s}_k = \pm \mathbf{e}_j, \quad \text{ha } k \equiv j \pmod{n},$$

- Véletlenszerűen
- Legnagyobb csökkenési irány:  $s_k = -\nabla f(x_k)$ .
- Newton-szerű módszerek:  $s_k = -H_k^{-1} \nabla f(x_k)$ , ahol  $H_k$  szimmetrikus pozitív definit mátrix.
  - Lederiváljuk minden változót által és behelyettesítjük az adott pontot
  - Meghatározzuk a második deriváltakat (az első deriválás eredményét újra lederiváljuk) és meghatározzuk az  $x_1 x_2$  deriváltat is (először  $x_1$  szerint aztán  $x_2$  szerint deriválunk) és mint előbb behelyettesítjük a megadott pontot
  - Második deriváltakat berakjuk Hesse mátrixba, előbb csak a  $x_1 x_2$  szerinti deriváltat határoztuk meg de ez egyenlő a  $x_2 x_1$  szerinti deriváltal
  - A hasse mátrixra és az első derivált behelyettesítése után kapott vektorral elvégezzük a newton módszert

T7. Az  $f(x) = x_1^3 + 2x_2^2$  függvény esetén az  $(1; 1)$  pontbeli Newton-lépések:

## Eso desivaltah:

$$\frac{dx}{dx_1} = 3x_1^2 \quad \frac{dx}{dx_2} = 4x_2$$

$\triangleright f(x) = \begin{pmatrix} 3x_1^2 \\ x_2 \end{pmatrix}$  es  $(1,1)$  punto ban  $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$

Násobit deriváta h ( $x_1 x_1 \rightarrow x_2 x_2 \rightarrow x_3 x_3$ )

$$\frac{\partial}{\partial x_1}x_1 = 6x \quad \frac{\partial}{\partial x_2}x_2 = 4 \quad \Rightarrow (1,1) = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

$$\frac{d}{dx_1 x_2} = 0 \quad (x_1 x_2)^1 = (x_2 x_1)^1$$

## flexe matrix

$$H \begin{pmatrix} x_1 x_1 & x_1 x_2 \\ x_2 x_1 & x_2 x_2 \end{pmatrix}$$

$$\text{NEWTON} : -H_{\ell}^{-1} \cdot \nabla f(x_{\ell})$$

$$-\begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = -\begin{pmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{5} \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$$

$\uparrow$                      $\uparrow$   
 $+ \quad \quad \quad \nabla f(x)$

- Az eredeti Newton-módszer konkrétan:

$$H_k = \nabla^2 f(x_k) (\text{ es } \alpha_k = 1)$$

- Hooke-Jeeves-módszer:

- $s_k$  keresési irányok a  $\pm e_1, \pm e_2, \dots, \pm e_n \in R^n$  egységvektorok ciklikusan, azaz

$$s_k = \pm e_j, \text{ ha } k \equiv j \bmod n$$

- o a  $\pm$  közül minden esetben az, amelyik csökkenési irányba mutat. (A konvergencia nem minden esetben)

garantálható.)

- A learning rate megválasztása
  - Konstans.
  - Heurisztikus, pl.:  $\alpha_k = \frac{1}{k+1}$
  - Minimalizáló:  $\alpha_k = \arg \min_{\alpha \geq 0} f(x_k + \alpha s_k)$
- Kombinált módszerek
  - Newton-módszer vonalmenti minimalizálással.
  - DFP (Davidon–Fletcher–Powell) algoritmus:
    - BFGS algoritmus vonalmenti minimalizálással

## LINEÁRIS REGRESSZIÓ

---

## Ismétés és kiegészítés

- Adottak az  $(x_i; y_i), i = 1, 2, \dots, n$  megfigyelések  $(X; Y)$ -ból.
  - Cél megtalálni az  $f(Y|X)$  regressziós függvény becslését, amely minimalizálja  $E * ((Y - f(X))^2) - t$
  - Tapasztalati úton: keressük  $\arg \min_f \frac{1}{n} \sum_i (y_i - f(x_i))^2$ -t egy nem túl nagy függvényosztályból ( $f \in F$ )
  - például a nyers leíró változók lineáris kombinációi közül:
$$f(x_i) = \beta_0 + \sum_{j=1}^P \beta_j x_{i;j}$$
- Ha az  $F$  függvényosztály véges dimenziós lineáris tér (pl. a legfeljebb d fokú polinomok), akkor f becsülhető az  $f_1, \dots, f_r$  bázis rögzítésével. Ekkor
 
$$\hat{f} = \sum_k f_k \hat{\beta}_k$$
, ahol
 
$$\hat{\beta}_k = \arg \min_{\beta} \frac{1}{n} \sum_i (y_i - \sum_k f_k(x_i) \beta_k)^2.$$
  - Tehát a függvényosztályon végzett regresszió (r dimenziós) lineáris regresszióvá egyszerűsödik a transzformált adatokon:  $(f_1(x_i); \dots; f_r(x_i); y_i)$ .
- Geometriai szemlélet:
  - $y \in R^n$  a célváltozó (response) vektor,  $X$  pedig a leírók (predictor), konstanssal kibovített mátrixa, a tapasztalati hiba (empirical risk) pedig
 
$$R_{emp}(\beta) = \frac{1}{n} \| y - X\beta \|^2$$
  - Az  $R_{emp}$  minimalizálásához először  $y$ -t az  $X$  oszlopterébe vetítjük
    - ez lesz  $\hat{y}$ .
  - Majd  $\hat{y}$ -ot kifejezzük  $X$  oszlopvektorainak lineáris kombinációjaként
    - így  $\hat{\beta}$  komponensei lesznek az együttható

## Modellválasztás a lineáris modellek között

### Változók kiválasztása a lineáris modellben

- A modellek között a bennük szereplő változók tesznek különbséget. A cél, hogy megtaláljuk azt a modellt, amelynek legkisebb a várható teszt vesztesége:  $R(\beta) = E((Y - X\beta)^2)$ .
  - $R(\beta)$  másik neve: **risk**. A  $\hat{\beta}$  becslést az  $R_{emp}(\hat{\beta}) = \frac{1}{n} \| y - X\hat{\beta} \|^2$  tapasztalati risk minimalizálásával keressük.
- Minél bővebb a modell, annál kisebb az  $R_{emp}\hat{\beta}$  minimalizált tapasztalati risk.

- Vagyis gond, hogy az  $R_{emp}\hat{\beta}$  minimalizált tapasztalati risk alulbecsüli a valós risk értékét.

x

- Ha  $\beta^*$  az együtthatók a tényleges modellben, akkor

$$R_{emp}(\hat{\beta}) \leq R_{emp}(\beta^*), \text{ miközben}$$

$$E(R_{emp}(\beta^*)) = R(\beta^*)$$

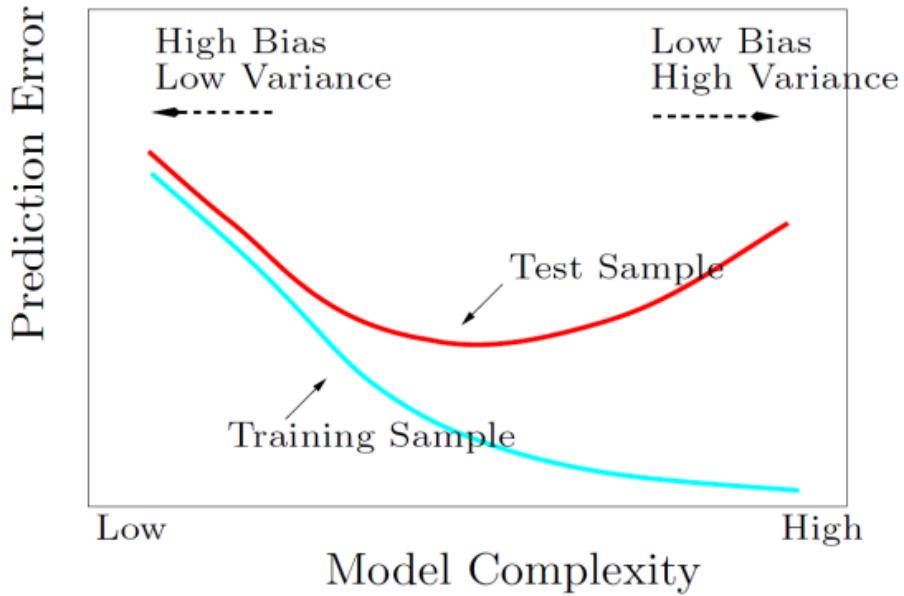
- Korrigálni szükséges a tapasztalati risk torzítását (bias). Általában használt módszerek:

- Mallow-féle  $C_p$ ,  $AIC$ ,  $BIC$  és a módosított  $R^2$ .
- Az illesztett modellek a fenti értékek alapján hasonlítjuk össze, és a legkisebb  $C_p$ ,  $AIC$  vagy  $BIC$  értékű, illetve a legnagyobb módosított  $R^2$  értékű modellt választjuk.

## A Teszthiba torzítás-bizonytalanság felbontása

- $Y = f(X) + \varepsilon$ , ahol  $\varepsilon$  (fehér) zaj (szórásnégyzete  $\sigma^2$ ),  $\hat{f} \in F$  az  $f$  becslése az  $(x_i; y_i)$ ,  $(i = 1; \dots; n)$  tanító mintán. (Megj.:  $\hat{f}$  véletlenszerű, a tanító mintától függ.)
- A teszthiba várható értéke  $x_0$ -ban:
$$\begin{aligned} E((Y - \hat{f}(X)|X = x_0)^2) &= \sigma^2 + E((f(x_0) - \hat{f}(x_0))^2) = \\ &= \sigma^2 + f(x_0) - E(\hat{f}(x_0))^2 + D^2(\hat{f}(x_0)) \end{aligned}$$
  - Az első tag a zaj elkerülhetetlen bizonytalansága.
  - A második tag  $f$  becslésének négyzetes torzítása (squared bias).
  - A harmadik tag  $f$  becslésének bizonytalansága (szórásnégyzete, varianciája).
- Túl **egyszerű** modell: ( $F$  túl szűk halmaz) nem fedi fel az  $f$  valós összefüggést (underfitting).
- Túl **bonyolult** modell: ( $F$  túl bő halmaz) leköveti az  $\varepsilon$  zájt is (overfitting).
- A legkisebb teszthiba eléréséhez az éppen megfelelő komplexitású modellt kell megtalálni

- A továbbiak alulparaméterezett ( $p < n$ , leginkább  $p \ll n$ ) mintákra vonatkoznak  
**Torzítás-bizonytalanság kettőssége (bias-variance trade-off)**



## Modellválasztás

- Mallow-féle:**

- Az  $(x_i; y_i), i = 1, 2, \dots, n$ , adott minta esetén meg lehet becsülni a teszt hibáját, ha minden fontos leíró változó szerepel a modellben.
  - Ha valamelyik hiányzik, nagy lesz az **RSS** (residual sum of squares).
  - Egy korrekciós tagot kell alkalmazni, ha a kelleténél több változó szerepel a modellben.
- A teszthibát úgy becsüljük, hogy veszünk egy újabb „mérést”  $y_i$ -re, legyen ez

$$y'_i, i = 1, 2, \dots, n$$

- Belátható, hogy:

$$y' - \hat{y}^2 = \| y - \hat{y} \|^2 + 2d\hat{\sigma}^2$$

így a  $C_p$  statisztika:

$$C_p = RSS + 2d\hat{\sigma}^2$$

ahol:

- $d$  a modellbe bevont leíró változók száma ( $d$  dimenziós lineáris regresszió)
- $\hat{\sigma}^2$  pedig a zaj szórásnégyzetének becslése a legbovebb modellben.
- Azt a modellt választjuk, amelyre  $C_p$  minimális

- Akaike információs kritérium – AIC**

- Hasonló becslést kapunk a maximum likelihood becslés (MLE) módszerrel.

- A  $\beta$  paraméter becslését a negatív log-likelihood függvény minimalizálásával kapjuk.
- Adódik, hogy Gauss-zaj, azaz  $\varepsilon \sim N(0; \sigma^2)$  esetén azt a modellt kell választani, amelyre alábbi minimális.

$$AIC = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$$

- Megjegyzés: Más úton, de ugyanazt kaptuk,  $C_p$  és  $AIC$  lényegében ugyanaz, egymás számszorosai, ugyanott van a minimumhelyük

- **Bayes információs kritérium – BIC**

- A Bayes-statisztikai becslés módszerrel némileg más információs kritériumot kapunk. Az adódik, hogy Gauss-zaj esetén azt a lineáris modellt kell választani, amelyre alábbi minimális.

$$BIC = \frac{1}{n}(RSS + d\hat{\sigma}^2 \ln n)$$

- Megjegyzés: Mivel  $\ln n > 2$  bármely  $n > 7$  esetén, a  $BIC$  statisztika jobban „bünteti” a sok változót használó modelleket, mint  $C_p$  (vagy  $AIC$ ), így általában egyszerűbb (alacsonyabb komplexitású) modell választását eredményezi.

- **Módosított  $R^2$**

- Szintén népszerű eszköz a különbözo mennyiségek u leíró változót használó modellek közötti választásra a módosított  $R^2$  statisztika alapján történo választás. Emlékeztető:

$$R^2 = 1 - \frac{RSS}{TSS}, \text{ ahol}$$

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ és}$$

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2.$$

- Mivel újabb változók bevonásával  $RSS$  minden csökken, így  $R^2$  értéke noni fog. A módosított statisztika:

$$\text{Módosított } R^2 = 1 - \frac{RSS}{TSS} \cdot \frac{n-1}{n-d-1}$$

ahol  $d$  a modellbe bevont leíró változók száma.

- A  $C_p$ ,  $AIC$  és  $BIC$  statisztikákkal ellentétben itt azt a modellt kell választani, amelyre a módosított  $R^2$  maximális

- **Legjobb részhalmaz választás (best subset selection)**

- Ennek során az összes lehetséges változó kombinációt megvizsgáljuk, és kiválasztjuk azt a modellt, amelyik a legjobban illeszkedik az adatokra valamilyen előre meghatározott kritérium szerint.

- Legjobb részhalmaz kiválasztásának algoritmusa
  - minden  $d = 0, 1, \dots, p$  esetén vegyük az összes lehetséges,  $d$  változót használó modellt,  $\binom{p}{d}$  ilyen van.
    - Válasszuk ezek közül a szokásos módon a legkisebb tanuló hibájút (legkisebb  $RSS$  vagy legnagyobb  $R^2$ )
  - Vizsgáljuk a minden egyes  $d$ -re kiválasztott legjobb modellt,  $p + 1$  ilyen van.
    - Számítsuk ki mindegyikre a  $C_p$ ,  $AIC$ ,  $BIC$ , módosított  $R^2$  statisztikák valamelyikét.
  - Válasszuk ki azt a  $d$ -t, amelynek modellje a legjobban teljesít.
- Csak akkor kivitelezhető, ha  $p$  nem túl nagy, mivel az illesztendő modellek száma  $2^{p+1} - 1$

- **Lépésenkénti választás (stepwise selection)**

- ■ stepwise selection egy automatikus eljárás a regressziós modellek illesztésére, amely során a magyarázó változók kiválasztása automatikusan történik. minden lépésben egy változót adunk hozzá vagy vonunk el a magyarázó változók halmazából valamilyen előre meghatározott kritérium alapján
- Mohó algoritmusok, amelyek egymásba ágyazott modellsorozatot eredményeznek
- Előre haladó kiválasztás (forward stepwise selection):
  - Indulunk ki a csak konstanst tartalmazó modellbol:  $A_0 = \{0\}$ .
  - minden lépésben vegyük be a modellbe azt a változót, amelyik a legnagyobb változást eredményezi a tanuló hibán:  $A_{d+1} = A_d \cup \{\ell_d\}$ , ahol  $\ell_d \notin A_d$ , és amelyre  $RSS$  a legkisebb.
  - Folytassuk, amíg  $A_p$ -t nem kapjuk (amíg minden változót be nem választottunk).
- Visszafelé haladó kiválasztás (backward stepwise selection):
  - Indulunk ki a teljes modellbol: "  $A_p = \{0; 1; \dots; p\}$ ".
  - minden lépésben hagyjuk el a modellbol azt a változót, amelyik a legnagyobb változást eredményezi a tanuló hibán:  $A_d = A_{d+1} \setminus \{\ell_{d+1}\}$ , ahol  $\ell_{d+1} \in A_{d+1} \setminus \{0\}$ , és amelyre RSS a legkisebb.
  - Folytassuk, amíg  $A_0$ -t nem kapjuk (amíg a csak konstanst tartalmazó modellhez nem jutunk).
- A  $p + 1$  elemű  $A_0 \subset A_1 \subset \dots \subset A_p$  (egymásba ágyazott) modellsorozat elemei közül válasszunk  $C_p$ ,  $AIC$ ,  $BIC$  vagy módosított  $R^2$  szerint legjobbat

## Zsugor módszerek (shrinkage methods)

- A változók egyesével kiválasztása helyett lehetséges a null-modelltol (csak tengelymetszet, konstans illesztés) „folytonosan” eljutni a teljes modellig (az összes változót használó lineáris regresszió).
- A zsugor módszerek egy „bünteto” (regularizáló) tagot adnak a veszteséghez:

$$\sum_{i=n}^n (y_i - \beta_0 - \sum_{j=1}^p x_{i;j} \beta_j)^2 + \lambda \cdot J(\beta_1; \beta_2; \dots; \beta_p)$$

- Feltesszük, hogy  $\sum_{x=1}^n x_{i;j} = 0$  (a leírók centráltak). Ekkor  $\beta_0 = \bar{y}$ , és egyszerűbb  $y - \bar{y}$ -sal dolgozni.
- Ahhoz, hogy a regularizáció működjön, a leíró változóknak azonos skálán kell elhelyezkedni, azaz  $\frac{1}{n} \sum_{i=n}^n x_{i;j}^2 = 1$  minden leíró változó egységnyi szórású lesz, ha standardizáljuk:

$$\tilde{x} = \frac{x_{i;j}}{\sqrt{\frac{1}{n} \sum_{i=n}^n (x_{i;j} - \bar{x}_j)^2}}$$

- A  $\lambda$  finomhangoló paraméter (tuning parameter), és általában keresztsvalidációval határozzuk meg az értékét.

## Ridge Regresszió

- A leíró változók standardizáltak, a célváltozó centrált.
- $J(\beta) = \|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2$ , azaz:

$$\hat{\beta}_{ridge} = \arg \min_{\beta} \left( \sum_{i=1}^n (y_i - x_i \beta)^2 + \lambda \|\beta\|_2^2 \right)$$

ahol  $\hat{\beta}_{ridge}$  a **ridge regresszió** együtthatói.

- $\lambda = 0$  a teljes lineáris modellt eredményezi, míg  $\lambda \rightarrow \infty$  esetén az együtthatók 0-hoz tartanak.
- $\lambda$  növelésével a modell egyre kevésbé rugalmas, a torzítás nő, a bizonytalanság csökken.
- **NEM** eredményezi leíró változók tényleges kiválasztását

## Lasso Regresszió

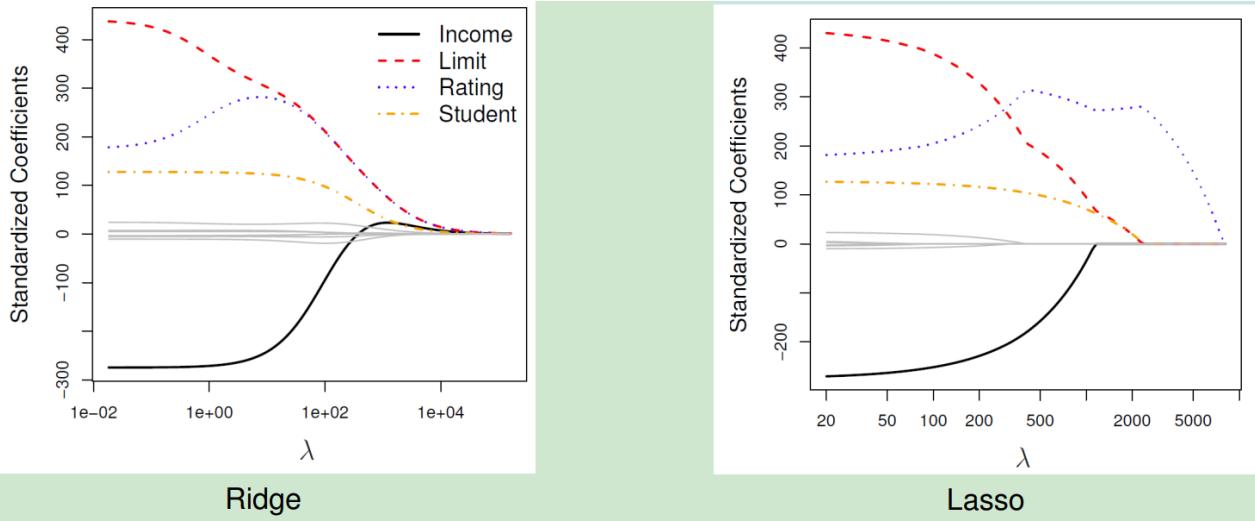
- A leíró változók standardizáltak, a célváltozó centrált.

$$\hat{\beta}_{lasso} = \arg \min_{\beta} \left( \sum_{i=1}^n (y_i - x_i \beta)^2 + \lambda \|\beta\|_1 \right)$$

ahol  $\hat{\beta}_{lasso}$  a lasso együtthatói.

- Nincs zárt alakban felírható megoldása
- $\lambda = 0$  a teljes lineáris modellt eredményezi
- míg  $\lambda \rightarrow \infty$  esetén az együtthatók 0-vá válnak.
- Leíró változók tényleges kiválasztását eredményezi

## Ridge & Lasso Regressziók Összehasonlítása



### Elastic net

- A ridge és a lasso lineáris kombinációja
  - NEM eredményezi leíró változók tényleges kiválasztását
- A leíró változók standardizáltak, a célváltozó centrált.
- $J(\beta) = \alpha \|\beta\|_2^2 + (1 - \alpha) \|\beta\|_1$ , ahol  $0 \leq \alpha \leq 1$ , azaz

$$\hat{\beta}_{elasticnet} = \arg \min_{\beta} \left( \sum_{i=1}^n (y_i - x_i \beta)^2 + \lambda \|\beta\|_2^2 + (1 - \alpha) \|\beta\|_1 \right)$$

ahol  $\hat{\beta}_{elasticnet}$  az elastic net együtthatói.

- $\alpha = 0$  a lasso
- míg  $\alpha = 1$  a ridge regresszióba megy át.
- A 2-es norma miatt a minimalizálandó függvény szigorúan konvex, így egyértelmű minimuma van

**this is shit - we are officially meth-ed up**

**FIN**

## FELADATTÍPUSOK / MEGOLDÁSOK

- 2. Hogyan írjuk fel egy lineáris leképezés mátrixát? Mekkora méretű lesz?
  - Bázisvektorokon kiszámítjuk a képet és azt berakjuk oszlopokban
  - 20d ből képzünk 4-be ( $R^{20x4}$ )
    - másik térben 4 komponensű oszlopvektorok és ezeket belerakjuk egy mátrixba
    - 4x20-as lesz az eredmény
    -
- 3.
 

Ismertesse a többváltozós (differenciálható) függvények szélsőértékelyének (analitikus) kiszámítási módját.

- gradiens vektor 0 vektor, azaz vízszintes az érintő síkja
- kell
- Shrinkage-módserek melyik normát használják? ( $\|\beta\|_2$ , vagy  $\|\beta\|_1$ )

**T10.** A zsugorító (shrinkage) módserek közül a 2-es normát használja regularizációra:

- (A) ridge és lasso      (B) ridge és elasticnet      (C) lasso és elasticnet      (D) csak a ridge

- Ridge: 2-es norma
- Lasso: 1-es norma
- Elastic: mindkettő
- Vonalmenti minumumkereső eljárások - Gradiens módszer

A gradiens módszert alkalmazzuk az  $f(\mathbf{x}) = x_1^2 - x_1x_2 + x_2^3$  függvényre az  $\mathbf{x}_0 = (1; 1)$  pontból indulva,  $\alpha = 0,1$  learning rate paramétert választva. Ekkor a következő iterációs pont

- (A)  $\mathbf{x}_1 = (0,9; 0,8)$       (B)  $\mathbf{x}_1 = (1,1; 1,2)$       (C)  $\mathbf{x}_1 = (0,8; 0,9)$       (D)  $\mathbf{x}_1 = (1,1; 1,1)$

$x_1$  derivált →  $\begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 3x_2^2 \end{bmatrix}$

$\nabla f(1,1) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$  → BEHELYERTETÉS

$-\nabla f(1,1) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$  → Feltő • -1

$-\alpha \nabla f(x_0) = \begin{bmatrix} -0,1 \\ -0,2 \end{bmatrix}$  → Felto II

- Megoldás : A, mert 1,1-hez hozzádtuk a  $[-0,1; -0,2]$  vektort
- Képtér

**T3.** Az  $\mathbf{A} \in \mathbb{R}^{4 \times 20}$  mátrixszal leírt lineáris leképezés a 20 dimenziós vektorokat \_\_\_\_\_ dimenziós képtérbe képezi.

- (A) 4      (B) legfeljebb 4      (C) 20      (D) legalább 4

### 3. Zérptérhez képest

$$\mathcal{L}: V \rightarrow V \quad \left. \begin{array}{c} \mathbb{R}^{4 \times 20} \\ \end{array} \right\} \mathcal{L} = \mathbb{R}^{4 \times 20}$$

Legfeljebb az „a” dim  
méretére képes az  $\mathbb{R}^{a \times b}$

- Mi lehet norma

T2. A  $\mathbf{v} = (x_1; \dots; x_n) \in \mathbb{R}^n$  vektorokon értelmezett alábbi függvények közül melyik norma?

- A  $\sum_{i=1}^n x_i$
- B  $\max_i \{x_i^2\}$
- C  $\sqrt[3]{\sum_{i=1}^n x_i^3}$
- D  $\sqrt[4]{\sum_{i=1}^n x_i^4}$

Norma általános alakja:

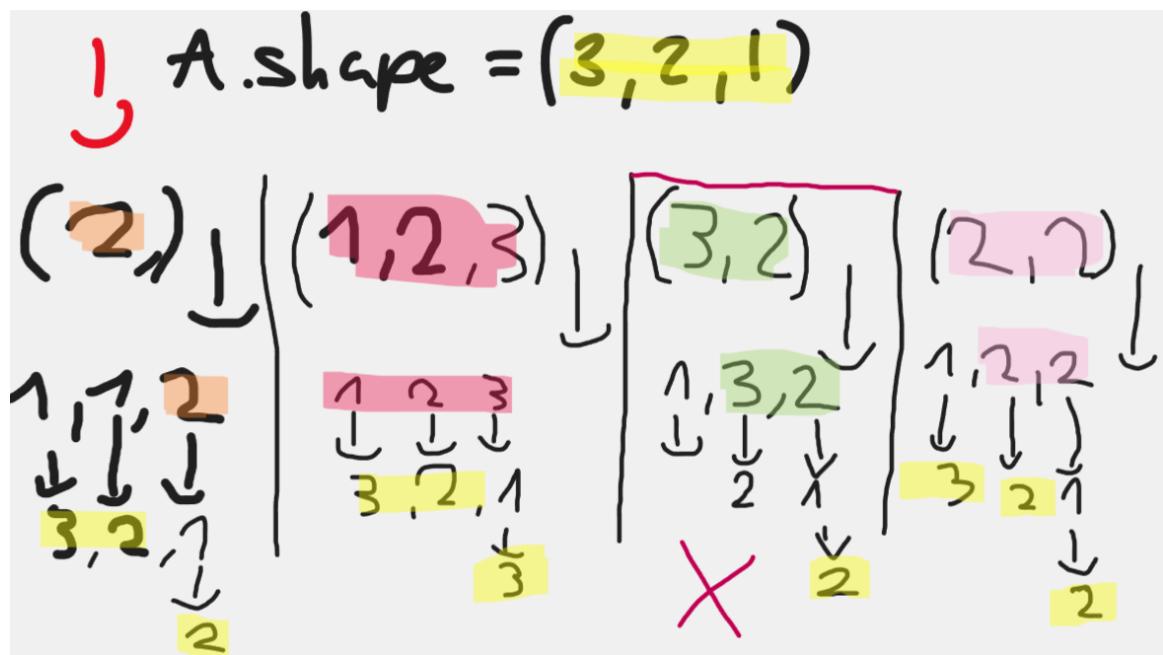
$$\sqrt{ \sum_{i=1}^n a_i^2 }$$

C  $\rightarrow$  nem jó, mert több, a mi több nem lehet c négyzetből rihozni

- Broadcasting

T1. Az alábbi méretű tömbök közül melyik összeadása nem értelmezett egy  $(3, 2, 1)$  méretű tömbbel?

- A (2,)
- B (1, 2, 3)
- C (3, 2)
- D (2, 2)



- SVD dimenzió

T5. Az  $A \in \mathbb{R}^{4 \times 20}$  mátrix  $U\Sigma V^T$  szinguláris érték felbontásában szereplő mátrixok mérete

- A  $4 \times 4, \quad 4 \times 20, \quad 20 \times 20$
- B  $4 \times 20, \quad 20 \times 4, \quad 4 \times 20$
- C  $4 \times 4, \quad 4 \times 4, \quad 4 \times 20$
- D  $4 \times 20, \quad 20 \times 20, \quad 20 \times 20$

- $m \times r, r \times r, r \times n$  (mert  $n \times r$  transponáltja a  $r \times n$ )

- Mátrix definitség

T6. Az  $\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$  mátrix

$$2 \cdot (-5) - (-3) \cdot 5 = -10 + 15 = 5 \neq 0$$

- A negatív definit.
- B indefinit.
- C pozitív definit.
- D negatív semidefinit.

$A = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$  definitsége

$$A_1 = [2] \quad A_2 = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$$

- $\det A_1 = 2 > 0$
- $\det A_2 = (2 \cdot 5) - (-3 \cdot -3) = 10 - 9 = 1 > 0$

$$A = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix} \left\{ \begin{array}{l} \det [2] = 2 > 0 \\ \det [A] = 10 - 9 = 1 > 0 \end{array} \right. = \phi$$

- Norma számolás:

2. Mennyi  $\|(6; -2; -3)\|_1$ ? ~~11~~

- A 1
- B 6
- C 7

2)  $\|(6; -2; -3)\|_1$ ,  
 norma átl. elv szerint  $\sqrt{\sum_{i=1}^3 n_i^2} \rightarrow \|v\|_2$   
 $\sqrt{6^2 + (-2)^2 + (-3)^2} = \sqrt{36 + 4 + 9} = 7$

Manhattan norma:  $\sum |a_i| \rightarrow \|v\|_1$   
 $6 + 2 + 3 = 11$

Pálosi Ákos

Sebíser norma:  $\max\{|x_1|, |x_2|, \dots, |x_n|\} \rightarrow \|v\|_\infty$   
 $\max\{6, 2, 3\} = 6$

○ Lásd Norma (<https://usnotes.szerver.cc/s/pRiVgvTOu#NORMA>)

- sajátérték és sajátvektor számításra

6. Az  $\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$  mátrix sajátértékei

- C pozitívak.
- D negatívak.
- B között van pozitív és negatív is.
- D között van nulla.

6)

$$\begin{aligned}\varphi(s) &= \lambda s, \\ As &= \lambda s, \\ As - \lambda s &= 0, \\ (A - \lambda E)s &= 0, \\ \det(A - \lambda E) &= 0.\end{aligned}$$

$$A = \begin{bmatrix} 2 & -5 \\ -3 & 5 \end{bmatrix}$$

$$\lambda E = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$A - \lambda E = \begin{bmatrix} 2 & -5 \\ -3 & 5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \quad \rightsquigarrow$$

- $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$

$$\det \begin{pmatrix} 2-\lambda & -5 \\ -3 & 5-\lambda \end{pmatrix} = \{(2-\lambda)(5-\lambda)\} - \{-3 \cdot (-3)\} =$$

$$10 + \lambda^2 - 5\lambda - \lambda^2 =$$

$$-10 + \lambda^2 - 5\lambda + 9 = \underline{\underline{\lambda^2 - 5\lambda + 1}}$$

$$\lambda^2 - 5\lambda + 1 = 0 \quad \boxed{\lambda_1 = 6,85} \quad \boxed{\lambda_2 = 0,14} \quad \leftarrow \text{rajz A elöljeléhez}$$

↓

$$\boxed{\begin{aligned} C_1 &= \sqrt{6,85} \\ C_2 &= \sqrt{0,14} \end{aligned}}$$

- stacionárius pont meghatározása

-  alábbiak közül melyik az  $f(x) = 8x_1 - x_1^2 x_2 + x_2^2$  függvény stacionárius pontja?
  - (A) (0; 0)
  -  (2; 2)
  -  (4; 8)
  - (D) (1; 4)

7)  $f(x) = 8x_1 - x_1^2 x_2 + x_2^2$

$$\frac{\partial}{\partial x_1} = 8 - 2x_1 x_2 \rightsquigarrow 8 - 2x_1 y = 0 \Rightarrow$$

$$\frac{\partial}{\partial x_2} = -x_1^2 + 2x_2 \rightsquigarrow -x_1^2 + 2y = 0$$

$$\begin{cases} 8 - 2x_1 y = 0 \\ -x_1^2 + 2y = 0 \end{cases} \quad \begin{aligned} 2x_1 y &= 8/2 \\ x_1 &= 4 \end{aligned}$$

$P(2,2)$ ,  $P_c(1,4)$ ,  $P_a(4,1)$

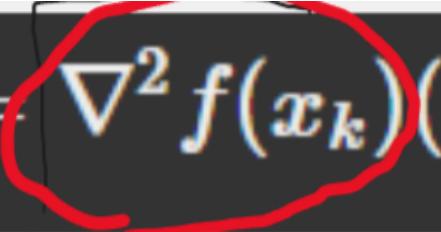
$$-x_1^2 + 2y = 0$$

$P_o(0,0)$ ,  $P(2,2)$  STAC

- Mi kell newtonhoz?

8. A Newton-módszeren alapuló numerikus minimumkereső eljárásban az  $x_{k+1}$  iterációs pont kiszámításhoz szükség  van az  $x_k$  pontbeli gradiensre.  Hesse-mátrixra.  minden kettőre.  egyikre sem.

$H_k = \nabla^2 f(x_k)$  ( és  $\alpha_k = 1$ )





GRADIA'NS VÉKTOR

- Newton lépés

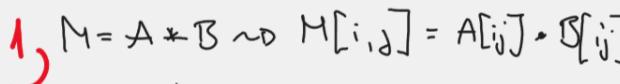
18. A Newton-módszeren alapuló numerikus minimumkereső eljárásban egy Newton-lépés:

- A  $\nabla f(x)$   B  $-\nabla f(x)$   C  $[\nabla^2 f(x)]^{-1} \nabla f(x)$   D  $-[\nabla^2 f(x)]^{-1} \nabla f(x)$

- Lásd Teszthiba (<https://usnotes.szerver.cc/s/pRiVgvTOu#A-Teszthiba-torz%C3%ADt%C3%A1s-bizonytalans%C3%A1g-felbont%C3%A1sa>)

- dot product

1. Az alábbi műveletek közül melyik nem minden esetben értelmezett két azonos méretű A és B tömb között?  A  $A * B$   B  $A ** B$   C  $A @ B$   D  $A - B$



Szorzás: ha  $A.shape == (m,n)$  és  $B.shape == (n,p)$ , akkor  $M = A @ B$  értelmezett,  $M.shape == (m,p)$ , és  $M[i,j] = (A[i,:]*B[:,j]).sum()$ .  
 $M = A.dot(B)$  vagy  $M = A @ B$

$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$   $B = \begin{bmatrix} 2 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 \end{bmatrix}$   $A.shape = (m,n) = (4,2)$   
 $B.shape = (n,p) = (2,4)$

$A @ B$

$M[1,1] = A[1,:]*B[:,1].sum()$   $M.shape = (4,4)$

$M[1,1] = (1,2) \cdot (1,2) = (1,2) \cdot 2$

- Lásd alapok (<https://usnotes.szerver.cc/s/pRiVgvTOu#1-alapfogalmak-%C3%A9s-m%C5%B1veletek>)

- SVD sajátétek

- 5. Az  $A \in \mathbb{R}^{4 \times 20}$  mátrix  $U_1 \Sigma_1 V_1^T$  redukált szinguláris érték felbontásában szereplő  $\Sigma_1$  mátrix mérete biztosan nem lehet
  - (A)  $2 \times 2$ .
  - (B)  $3 \times 3$ .
  - (C)  $4 \times 4$ .
  - (D)  $5 \times 5$ .

$R^{4 \times 20} \rightsquigarrow$  legfeljebb  $4 \times 4$

- 4. A valós mátrixok szinguláris értékei
  - (A) nem-negatívak.
  - (B) különbözők.
  - (C) nem nullák.
  - (D) nem minden pozitív.

■ A szinguláris értékek a mátrix sajátértékeinek pozitív négyzetgyökei

#### T4. A valós szimmetrikus mátrixok sajátértékei

- (A) pozitívak, ha ... ? (B) különbözők.
- (C) nem nullák.
- (D) valósak.

- Mallow:

- 1.:

- 9. A Mallow-féle  $C_p$  (AIC, BIC, stb)
  - (A) a tanulási hibát (train error) becsüli.
  - (B) a teszt hibát (test error) becsüli.
  - (C) a teszt hiba torzítását (bias) becsüli.
  - (D) a teszt hiba bizonytalanságát (variance) becsüli.
- Az  $(x_i; y_i), i = 1, 2, \dots, n$ , adott minta esetén meg lehet becsülni a teszt hibáját, ha minden fontos leíró változó szerepel a modellben.
  - Ha valamelyik hiányzik, nagy lesz az RSS (residual sum of squares).
  - Egy korrekciós tagot kell alkalmazni, ha a kelleténél több változó szerepel a modellben.
- Lásd Teszthiba (<https://usnotes.szerver.cc/s/pRiVgvTOu#A-Teszthiba-torz%C3%ADt%C3%A1s-bizonytalans%C3%A1g-felbont%C3%A1sa>)

- 2.:

- T8. A Mallow-féle  $C_p$  (AIC, BIC, stb) azt a modellt választja, amelynek a legkisebb a
  - (A) teszt hibája.
  - (B) torzítása.
  - (C) komplexitása.
  - (D) bizonytalansága.
- (E) Mivel teszt hibát becsüli, this is obvious

- Ridge vagy lasso eredmény

- 10. A zsugorító (shrinkage) módszerek közül melyik eredményezi a leíró változók tényleges szelekcióját?
  - (A) csak a ridge
  - (B) csak a lasso
  - (C) ridge és lasso
  - (D) egyik sem

- A lasso a válasz mert a ridge regresszió nem eredményez leíró változók tényleges szelekcióját. A ridge regresszió célja, hogy csökkentse az együtthatók nagyságát, de nem teszi őket pontosan nullává. Ez azt jelenti, hogy a ridge regresszió nem választja ki a változókat, hanem csak csökkenti azok hatását a modellben
- Lásd Zsugór módszerek (<https://usnotes.szerver.cc/s/pRiVgvTOu#Zsugor-m%C3%B3dszerek-shrinkage-methods>)
-

- Metrika

- 3. Melyik tulajdonság nem jellemző egy általános metrikára?
  - (A) nem-negatitás
  - (B) szimmetria
  - (C) skálazhatóság
  - (D)  $\triangle$ -egyenlőtlenség
- A skálazhatóság nem jellemző tulajdonsága az általános metrikának, negatív nem lehet a gyök miatt a másik kettő meg tulajdonsága Lásd Metrika  
(<https://usnotes.szerver.cc/s/pRiVgvTOu#METRIKA>)

- Komplex modellek

T9. Melyik nem jellemző a túl sok változót használó regressziós modellre?

- (A) alacsony tanulási hiba (train error)
- (B) nagy teszt hiba (test error)
- (C) nagy torzítás (bias) a teszt hibában
- (D) nagy bizonytalanság (variance) a teszt hibában
- Lásd Teszhiba (<https://usnotes.szerver.cc/s/pRiVgvTOu#A-Teszthiba-torz%C3%ADt%C3%A1s-bizonytalans%C3%A1g-felbont%C3%A1sa>)

- Tömb indexelés

- Minden helyes válasz 1 pont. A szöveges válaszoknál törtörtezzetek.
 

*Minden helyes válasz 1 pont. A szöveges válaszoknál törtörtezzetek.*

*T1. Az M tömb M[-1,0] eleme* *1,0,-1,0,-1*

*(A) a jobb felső.* *a bal alsó.* *C) a jobb alsó.* *D) egyik sem.*
- *1)  $\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix} = M$*

*M[-1,0] ~> utolsó sor*  
*∅. eleme*

- Mátrix műveletek

- T2. Az A tömb (1,3), míg a B tömb (3,2,1) méretű. Az alábbi műveletek eredménye közül melyik mérete különbözik a másik háromtól?
  - (A)  $A + B$
  - (B)  $A * B$
  - (C)  $A @ B$
  - (D)  $B @ A$

*nen végzhető* *nen végzhető el*

$\text{A} + \text{B}$        $M = A * B$   
 $(1,3) + (3,2,1)$        $M[i,j] = A[i,j] \cdot B[i,j]$   
 ○  $3+1 \checkmark \approx 5+3$   
 $1+2 \checkmark \approx 2+2$        $(3,2,1)$   
 $1+3 \checkmark \approx (1,3) = (1,1,3) \approx 1+3 \checkmark$   
 $\downarrow \qquad \qquad \qquad \begin{matrix} 2 \\ 3 \end{matrix} \qquad 3+3$   
 $(3,2,1) + (3,2,3)$

- $\|u\|_1$  és  $\|v\|_2$  közti kapcsolat

T3. A  $\|v\|_1 \geq \|v\|_2$  egyenlőtlenség  
 ○  minden  $v$  vektorra igaz.    
 ○  csak a nullvektorra igaz.   
 ○  egy  $v$  vektorra sem igaz.  csak a bázisvektorok számszorosára igaz.

$\text{3) } \|v\|_1 = |x_1| + |x_2|$   
 $\|v\|_2 = \sqrt{x_1^2 + x_2^2}$   
 $v = [x_1 = 1; x_2 = 2]$

○  $P_1 \quad v_1 = 1+2=3$   
 $\|v\|_2 = \sqrt{1+4} = \sqrt{5} \approx 2,236$   
 $v = [x_1 = 2; x_2 = 3]$

$P_2 \quad \|v\|_1 = 5$   
 $\|v\|_2 = \sqrt{4+9} = \sqrt{13} \approx 3,6$

- Szinguláris érték

- T5. Az alábbi mátrixok közül melyiknek van 0 szingularis értéke?
- A  $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$
  - B  $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$
  - C  $\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ 0 & 6 \end{bmatrix}$
  - D  $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 0 \\ 1 & 6 & 0 \end{bmatrix}$

5)  $\det = \sqrt{\text{scj. zrt}}$

$$\det(M - \lambda E)$$

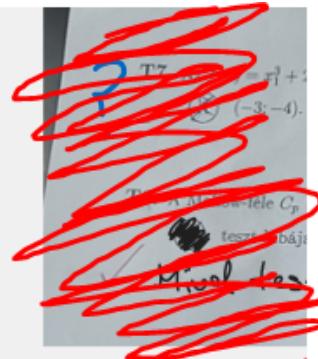
$$\det(A) = (1 \cdot 6) - (2 \cdot 3) = 0$$

o  $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1-7 & 2 \\ 3 & 6-7 \end{bmatrix} = \det(M) - (1-7)(6-7) - 6 =$$

$$6 - 1 - 67 + 7^2 - 6 = 7^2 - 77 \approx 7^2 - 77 + 0 = 0 \quad \begin{cases} \lambda_1 = 7 \\ \lambda_2 = 0 \end{cases}$$

SVD értékek =  $\sqrt{7}$  és a 0



- o Ha egy mátrix determinánsa 0, akkor az **egyik sajátértéke** is 0. Ez azért van, mert a mátrix sajátértékeinek szorzata megegyezik a mátrix determinánsával
  - Mivel a Szinguláris érték a sajétértek négyzetgyöke(definíció), ezért egyértelműen 0 lesz az egyik sajátértéke (ahogy az le is lett vezetve)

- Többváltozós  $f(x)$  tulajdonsága

T6. Az  $f(x)$  többváltozós függvény  $x_0$  pontbeli második deríváltja (Hesse-mátrixa), azaz  $\nabla^2 f(x_0)$  pozitív definit.

- o Ekkor az  $x_0$  pontban az  $f$  függvény
  - A -nek minimuma van.
  - B -nek maximuma van.
  - C konvex.
  - D konkáv.

- o A Hesse-mátrix szimmetrikus, szerencsés esetben ( $f$  szigorúan konvex) pozitív definit

- 2 és 3 dimenziós mátrixösszeszorzása ( $A @ B$ )

- o Milyen feltételekkel kell teljesülni, hogy egy 2- és egy 3-dimenziós tömb összeszorozható legyen (mátrixszorzás, azaz @ szerint)? Hány dimenziós és milyen méretű lesz az eredmény?

1)  $\begin{cases} A.shape = (m, n) \\ B.shape = (n, p) \end{cases} \quad M = A @ B \text{ új } M.shape = (m, p)$

- o -  $M = A @ B$  esetén  $M$  az  $A[i]$ -t és a  $B[j]$  shape-est veszi fel

- amennyi oszlopai van  $A$ -nak  
- annyi sorai legyen  $B$ -nak

- o Ez 3+ demenzió is működik:

- When calculating the dot product( $A @ B$ ) of two N-D arrays  $A$  and  $B$  using `np.dot(A, B)`, the result is a sum product over the last axis of  $A$  and the second-to-

last axis of  $B$ . Specifically, if  $A$  has shape  $(i, j, k)$  and  $B$  has shape  $(k, l, m)$ , then the resulting array  $M$  will have shape  $(i, j, l, m)$

- $A.shape=(i,j,k)$ ,  $B.shape=(k,l,m)$  és  $M = A @ B$  akkor  $M.shape=(i,j,l,m)$
- A két szélső axis közreöleli a belső axis-eket

- Ortogonális mátrix

o **Mit jelent, hogy egy mátrix ortogonális? Mi az előnyös tulajdonsága, amit ki is használtunk?**

- o Az ortogonális mátrix, egy olyan négyzetes ( $n \times n$ -es) aminek a sor és oszlopvektorai ortogonálisak, azaz transponáltja megegyezik az inverzével  $Q^T = Q^{-1}$ 
  - Ha  $A$  ortogonális, akkor  $\det(A) = \pm 1$
  - Két ortogonális mátrix szorzata ortogonális lesz
  - Egy ortogonális mátrix inverze ortogonális
- o Megjegyzés/magyarázat:
  - When we say that the transpose of a matrix is equal to its inverse, we mean that if we take the transpose of the matrix and then multiply it by the original matrix, we get the identity matrix ( $E$ ).
  - **Transponált:** sor és oszlop cseréje -> The transpose of a matrix is obtained by flipping the matrix over its main diagonal. This means that the rows of the original matrix become the columns of the transposed matrix, and vice versa.
  - **Inverz/ $A^{-1}$ :** The inverse of a square matrix  $A$  is another square matrix  $A^{-1}$  such that  $A * A^{-1} = E$ , where  $E$  is the identity matrix. The identity matrix is a square matrix with 1s on its main diagonal and 0s everywhere else.
  - So, if an orthogonal matrix  $Q$  satisfies  $Q^T = Q^{-1}$ , then  $Q * Q^T = I$

- SVD felbontás értelme:

o **A mátrix szinguláris érték felbontásának elsődleges célja a dimenziócsökkenés. Mely dimenziókat hagyjuk el, és miért?**

- o Első körben azokat dobjuk ki, ahol a szinguláris értékeket 0-ák. Ezután nagyság szerint a kisebb szinguláris értékekhez tartozó dimenziókat hagyjuk el
  - (Tanicibá lehet csak annak a válasznak örül, ha a 0-át írja az ember. So use this on your own risk)
- o Az így kiválasztott elemek kevésbé fontosak az adatok variabilitásának megőrzése szempontjából. Ezáltal csökkenthető az adatok mérete anélkül, hogy jelentős információveszteség következne be

- Metrika, példával:

o **Mi a metrika? Mutasson rá egy példát is.**

- Metrika fogalma, és képlete:

**4. Általánosított vektortávolság**

$$\begin{aligned} \mathbf{x} &= [x_1; x_2] \\ \mathbf{y} &= [y_1; y_2] \end{aligned} \quad \left\{ d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \right.$$

- Pl:

- Vegyük  $x$  és  $y$  vektorokat, ahol  $[x_1 = 2; x_2 = 3]$  és  $[y_1 = 4; y_2 = 5]$
- Ilyenkor  $d(x, y) = \sqrt{(2 - 4)^2 + (3 - 5)^2}$  azt helő

- Minimumkereső eljárások (ez a rész úgy ahogyan van, benne van a jegyzetben) ()

~~fel a minimumkereső eljárások általános alakját. (Nevezze meg a benne szereplő paramétereiket.)~~

### Általános kereső eljárás

```

1 | input [a, b], ε > 0
2 | repeat
3 |   select c, d ∈ (a, b) such that c < d.
4 |   if f(c) < f(d)
5 |     b = d
6 |   else
7 |     a = c
8 |   end
9 | until b - a < ε

```

#### Középső pontok megválasztása:

- Az egyik belső pont az új intervallum végpontja lesz, a másik pedig belső pontja.
- Az utóbbi helyen már egyszer kiszámított függvényértéket jó lenne felhasználni a következő ciklusban, gyorsítva ezzel az eljárást.
- Kellene, hogy az új belső pont az új intervallumban is „jó” helyen legyen, vagyis a régi-új belső pont ugyanolyan arányban ossza az új intervallumot, mint a régit.
- **Megoldás: aranymetszés**

- középső pontok:  $c$  és  $d$ , ahol  $c < d$
- intervallum szélső értékei:  $a$  és  $b$ , ahol  $a < b$

- Newton lépés pontban

T7. Az  $f(\mathbf{x}) = x_1^3 + 2x_2^2$  függvény esetén az  $(1; 1)$  pontbeli Newton-lépés:

Első deriváltak?

$$\frac{dx}{dx_1} = 3x_1^2 \quad \frac{dx}{dx_2} = 4x_2$$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 3x_1^2 \\ 4x_2 \end{pmatrix} \text{ és } (1, 1) \text{ pontban } \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Második deriváltak ( $x_1x_1$  és  $x_2x_2$  és  $x_1x_2$ )

$$\frac{d}{dx_1x_1} = 6x \quad \frac{d}{dx_2x_2} = 4 \Rightarrow (1, 1) = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

$$\frac{d}{dx_1x_2} = 0 \quad (x_1x_2)^1 = (x_2x_1)^1$$

Jelölések

$$H \begin{pmatrix} x_1x_1 & x_1x_2 \\ x_2x_1 & x_2x_2 \end{pmatrix}$$

NEWTON :  $-H^{-1} \cdot \nabla f(x_0)$

$$-\begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1/6 & 0 \\ 0 & 1/4 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$$

$\uparrow \quad \uparrow$   
 $H \quad \nabla f(x)$

- Hány dimenziós a magtér

- T4. Hány dimenziós a  $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}^2, (v_1; v_2; v_3) \mapsto (v_1; v_2 + v_3)$  lineáris leképezés magtere?  $\dim \text{Ker}(\varphi) =$

A 0       B 1       C 2       D 3

- Az biztos hogy egy dimenziós az eredmény de hogy miért azt fasz se tudja, ebből ki lehet indulni de nem biztos hogy jó:
  - A  $\mathbb{R}^3 \rightarrow \mathbb{R}^2, (v_1; v_2; v_3) \mapsto (v_1; v_2 + v_3)$  lineáris leképezés magtere azon vektorok halmaza, amelyekre a leképezés értéke a nullvektor. Ebben az esetben a magteret az alábbi egyenletrendszerrel írhatjuk le:

$$\begin{aligned}v_1 &= 0 \\v_2 + v_3 &= 0\end{aligned}$$

- Ebből következik, hogy a magteret az  $(-v_3; v_3; v_3)$  alakú vektorok alkotják, ahol  $v_3 \in R$ . Ez azt jelenti, hogy a magteret egy egydimenziós altér alkotja, tehát a magterének dimenziója 1.

- Broadcasting

-  A *broadcasting* (kiterjesztés) elv a python programozási nyelvben. (Hol jelenik meg? Mire értelmezett? Hogyan értelmezett? Hogyan írható elő melyik dimenzió mentén legyen a broadcasting? Írjon példákat.)
-