

# ML és más állatfajták

CERN TWS

23.08.2018.

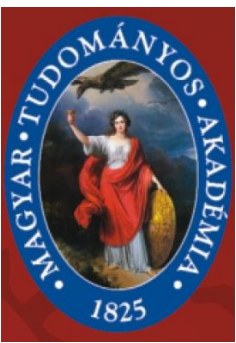
András Telcs

Acting head of Department of Computational Sciences.



NAP Pattern group

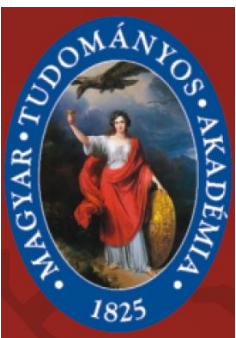
10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111



# Data everywhere!

---

1. **Google:** processes 24 peta bytes of data per day.
2. **Facebook:** 10 million photos uploaded every hour.
3. **Youtube:** 1 hour of video uploaded every second.
4. **Twitter:** 400 million tweets per day.
5. **Astronomy:** Satellite data is in hundreds of PB.
6. ...
7. **“By 2020 the digital universe will reach 44 zettabytes...”**



# ML



## Minták, tanulás, döntés

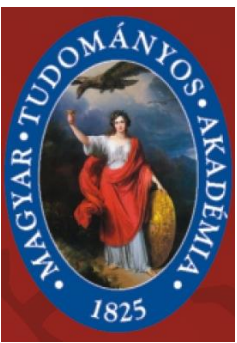
- Előzmény – diagnosztika – szakértői rendszerek
- Hitelbírálat
- Bank kártya csalás azonosítása – ügyfel profil
- Web profil - célzott hirdetés
- Arcfelismerés



HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

```
10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111
```



# ML



## Minták, tanulás, döntés

- Karakter felismerés
- Kézírás felismerés
- Beszédfelismerés



# Mi van a gépházban?

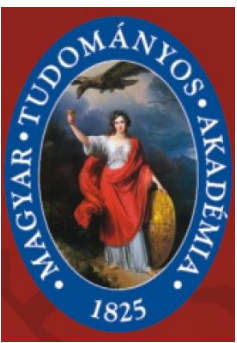
- Osztályozás
- Klaszterezés
- Előrejelzés
- Lényeg kiemelés

# Osztályozás

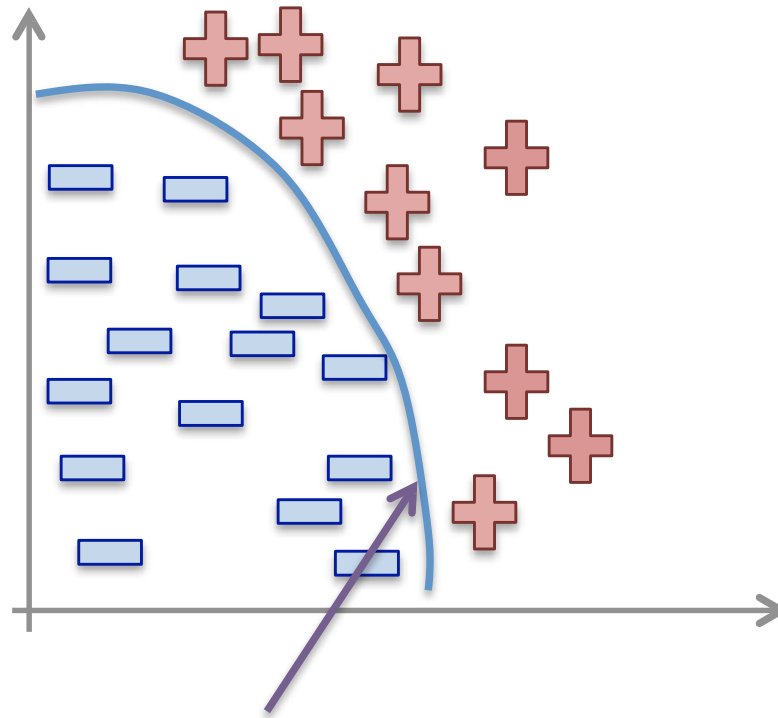
- Hitel (-elő)minősítés
- Karakter felismerés
- Diagnosztika

# Osztályozás

- Hitelminősítés
  - Jó adós
  - Rossz adós
- 0-1 döntés (nehéz)



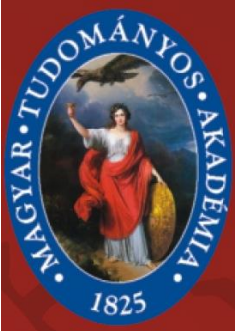
# Felügyelt tanulás



Módszerek: k-szomszéd, naiv Bayes, döntési fa, támaszvektor, Laplace módszer, neurális háló



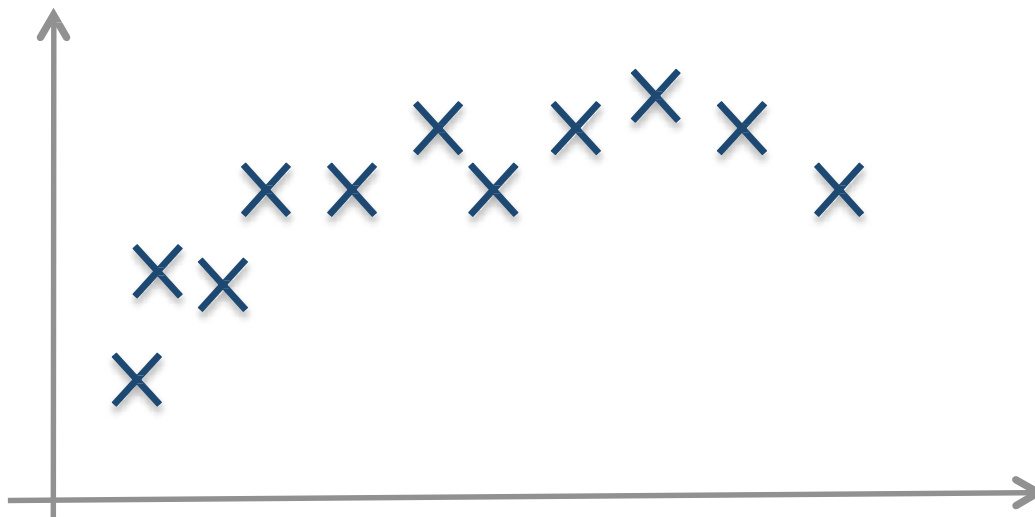




# Önálló tanulás



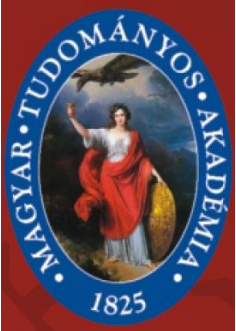
ős – regresszió  
Szabály tanulás



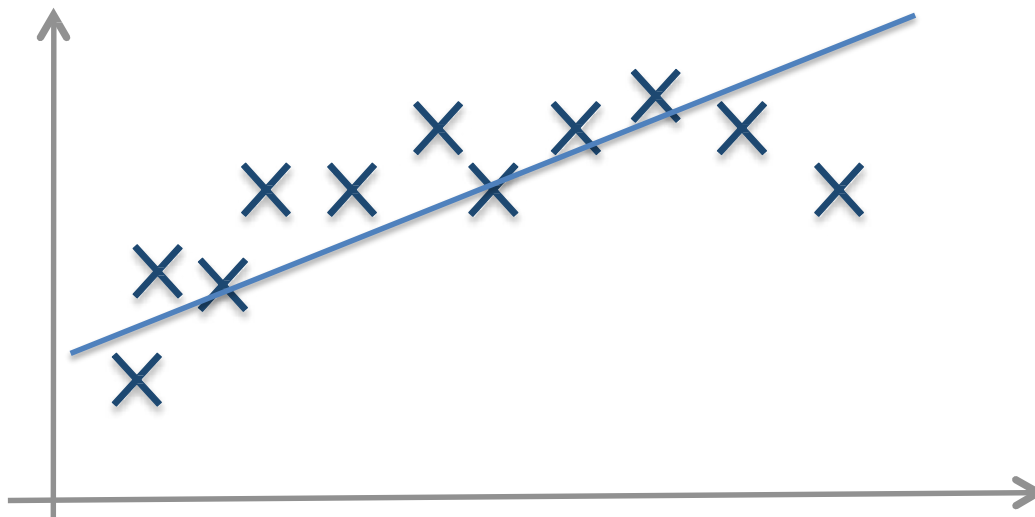
HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111



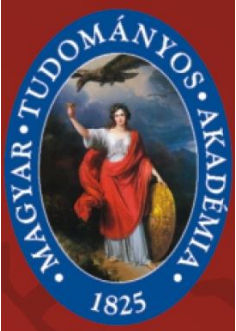
# Regresszió



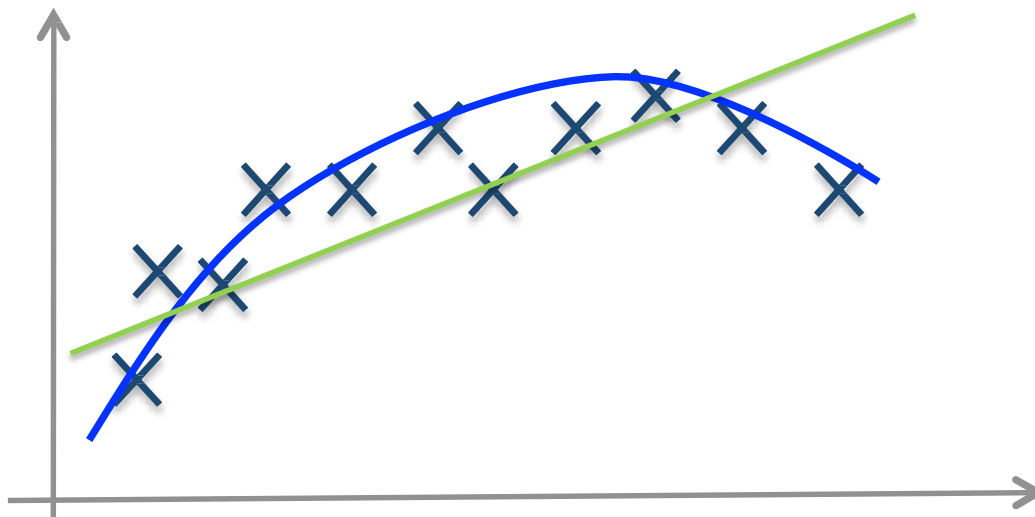
HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

```
10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111
```



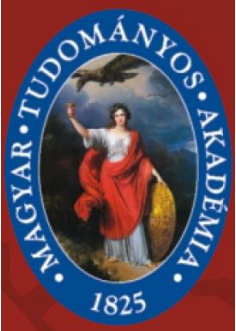
# Regresszió



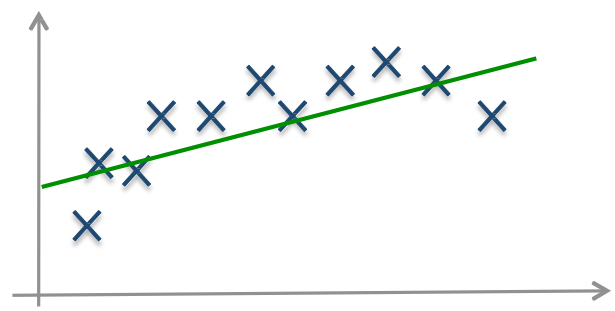
HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

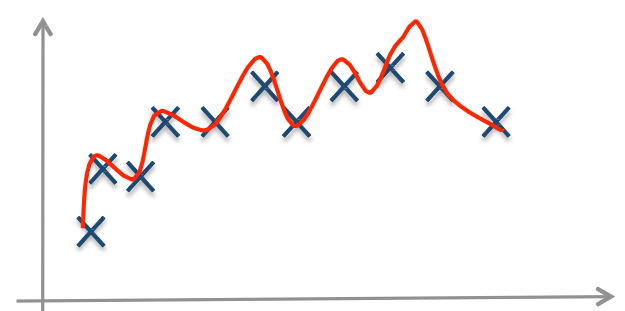
```
10P110100110111
00A011000001011
10T001010111001
01T100101110100
00E111001001101
00R010110000010
10N110100110111
```



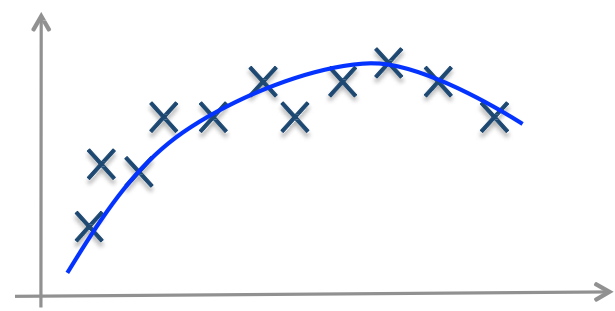
# Regresszió



Nagy torzítás  
Alul illesztés



Túl illesztés

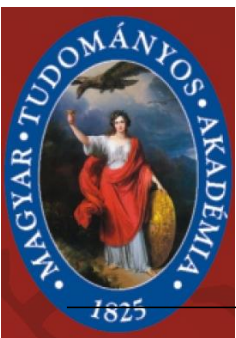


jó

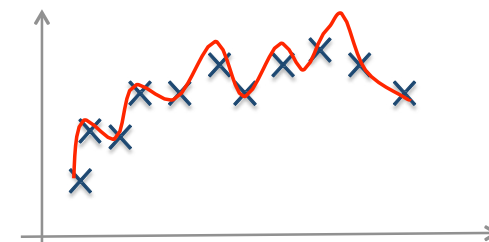
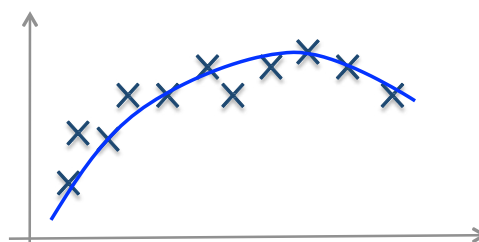
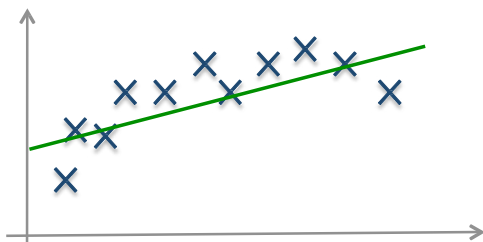


NAP Pattern group

```
10P110100110111
00A011000001011
10T001010111001
01T100101110100
00E111001001101
00R010110000010
10N110100110111
```



# Regularizáció:



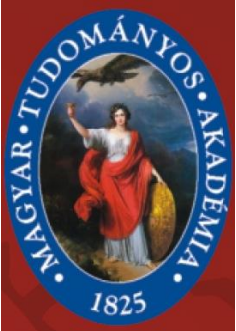
$$f(x) = \lambda_0 + \lambda_1 x \dots (1)$$

$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 \dots (2)$$

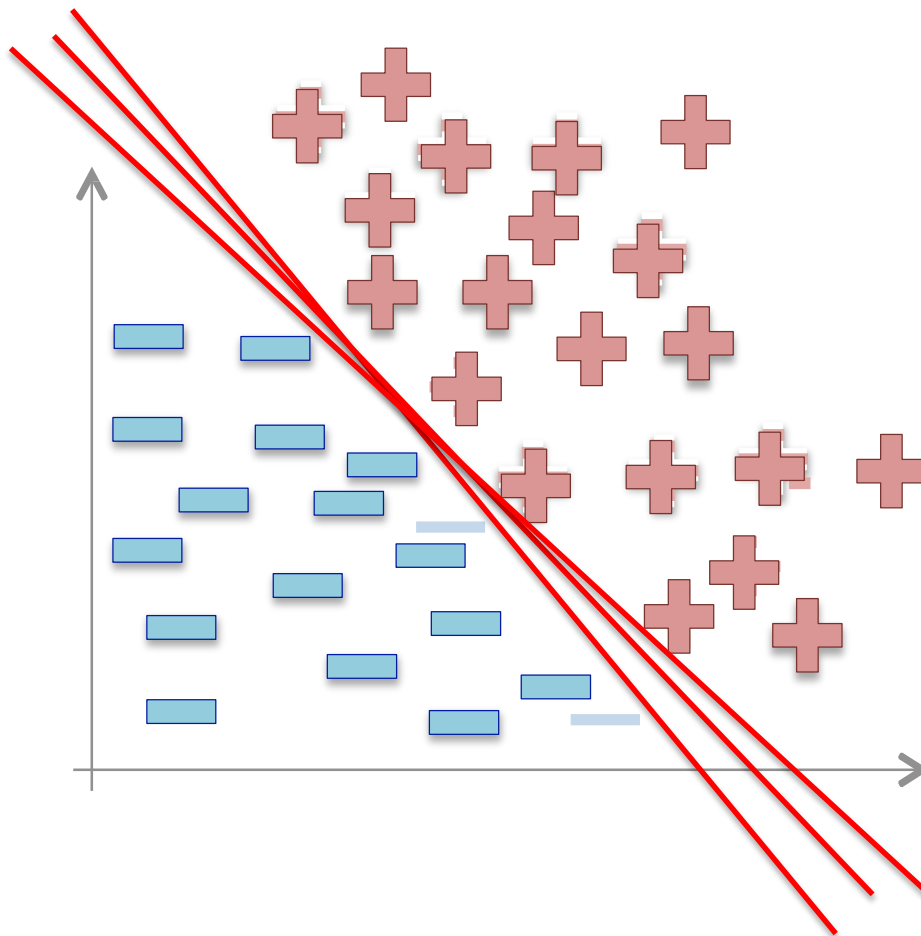
$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4 \dots (3)$$

Egyensúly a símaság és illeszkedés között





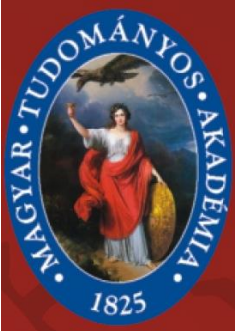
# Szeeparálás



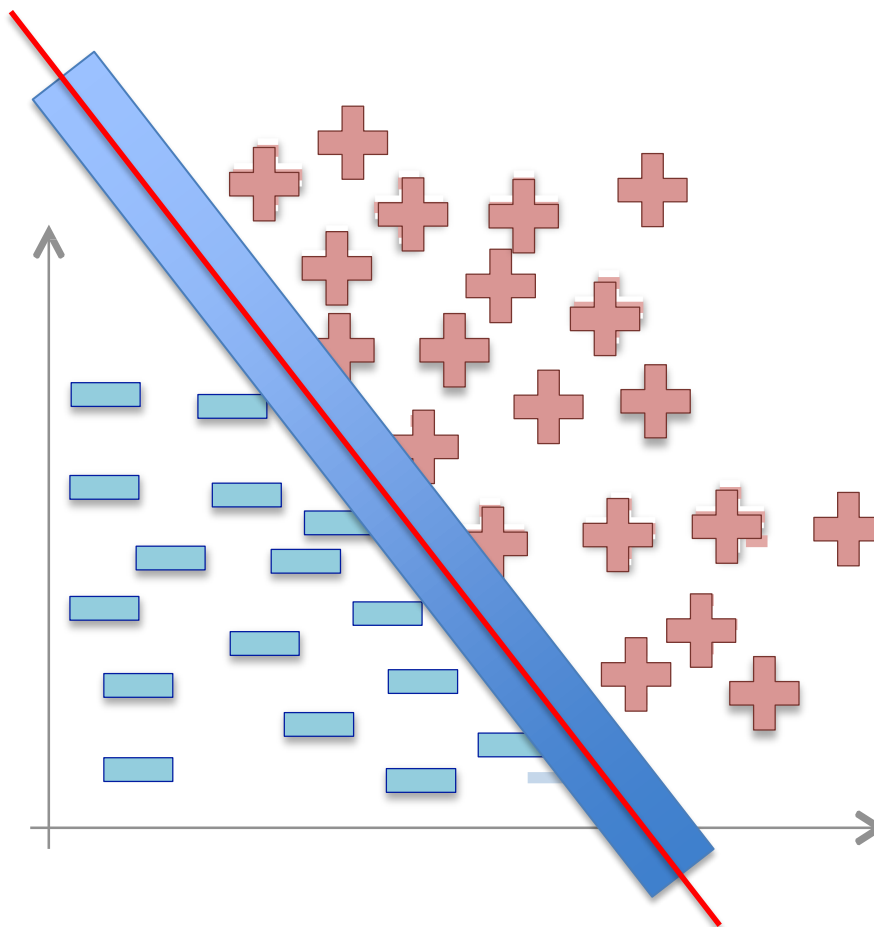
HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

```
10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111
```



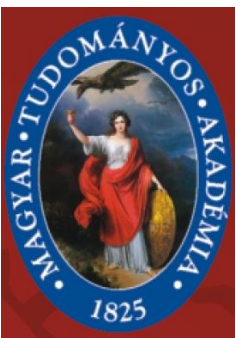
# Szeeparálás SVM



HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

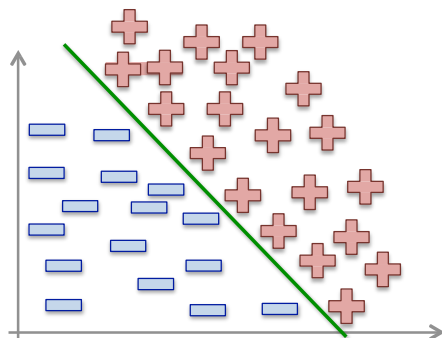
NAP Pattern group

```
10P110100110111
00A011000001011
10T001010111001
01T100101110100
00E111001001101
00R010110000010
10N110100110111
```

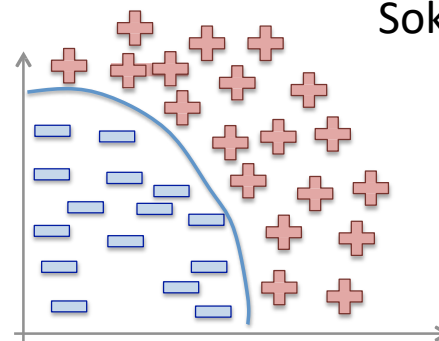


# ML alapeszközök

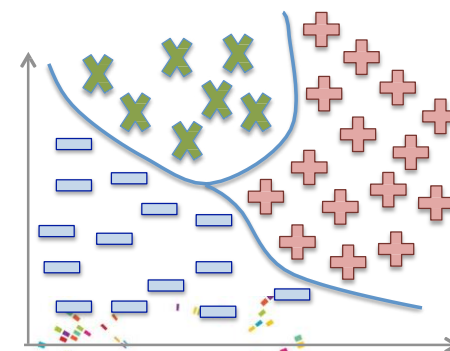
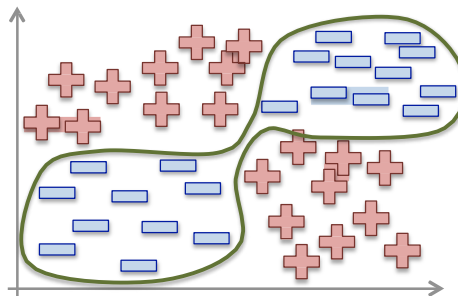
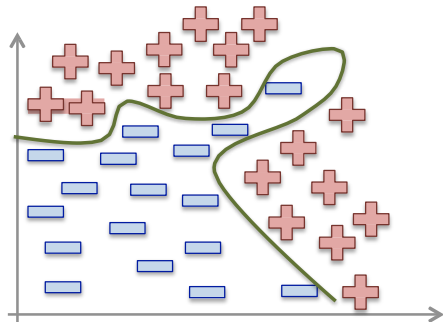
SVM



Sokaság tanulás

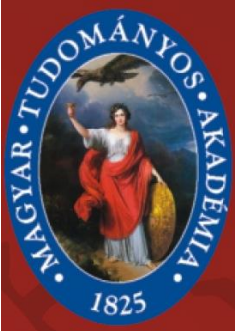


Osztályozás nem szokásos cilfüggvlnnyekkel



HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM



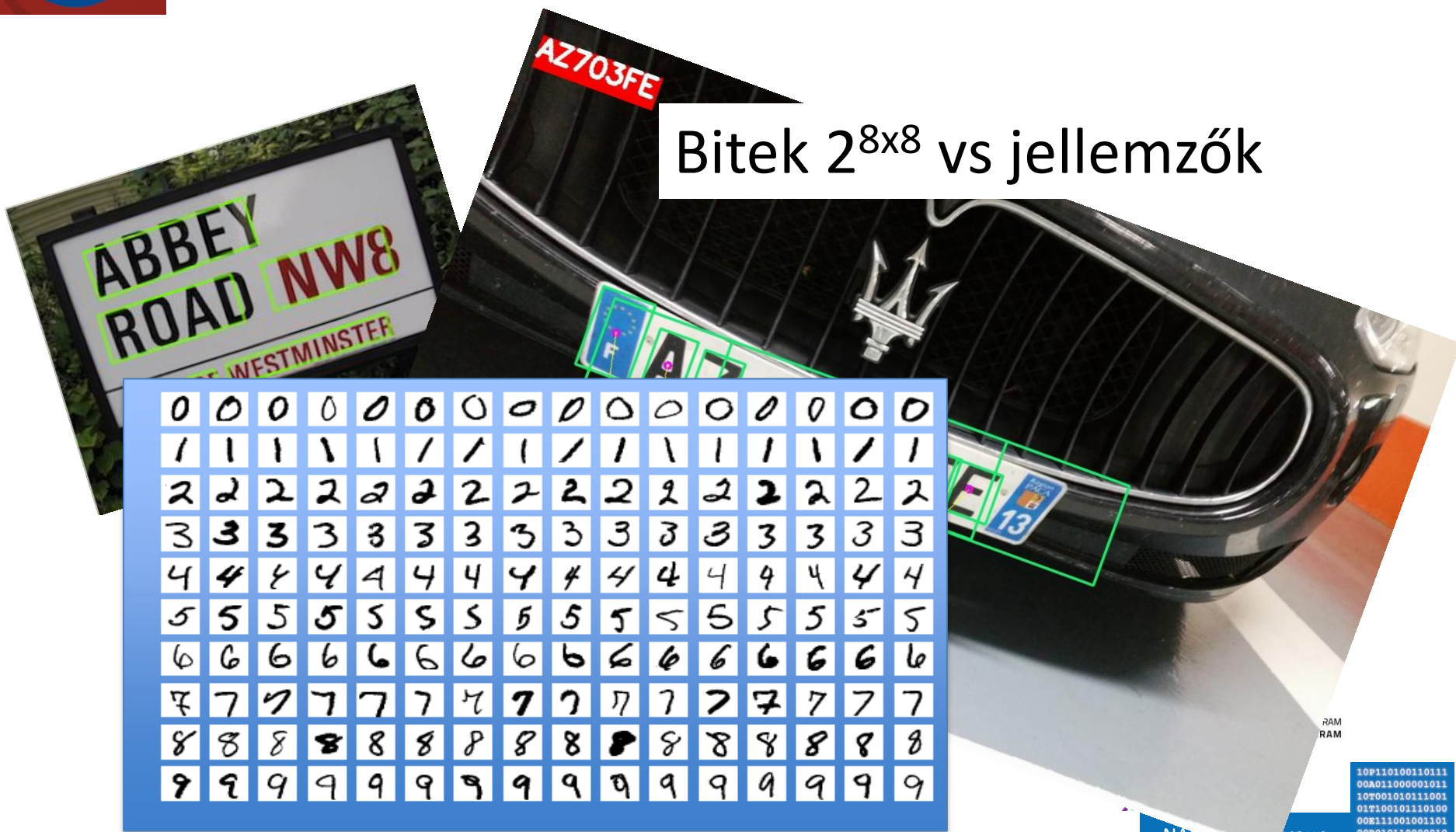


# Karakter felismerés

## jellemző kiemelés



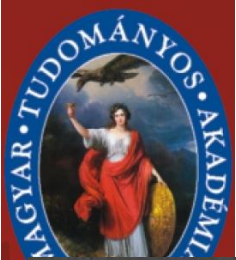
Bitek  $2^{8 \times 8}$  vs jellemzők



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111





jellemző kiemelés



# Arc felismerés

A nagy testvér figyel téged



HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

IAP Pattern group

10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111

# Személyiség profil

- Click
- Click
- Click

A nagy testvér figyel téged



# Google

- Web találatok sorbarendezése
- 10
  - hálózat weblinkekből
- 100
  - 5,000,000,000 =  $5 \times 10^9$  weblap
- 1000
  - $10^{12}$  web link
- 10,000
- 100,000 találat

$10^9 \times 10^9$  mátrix önmagával szorzása  
sokszor

# Csillagászat

$10^{11}$  galaxis

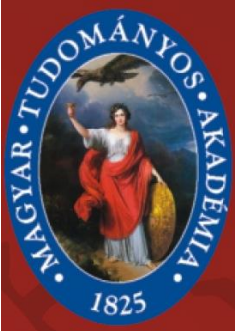
$10^{23}$  csillag

Feladat jellemzők szerint klaszterezés  
Spektrum, méret, alak, sűrűség eloszlás,  
távolság, elhelyezkedés....

Spektrum maga 1000+ tulajdonság  
(dimenzió)

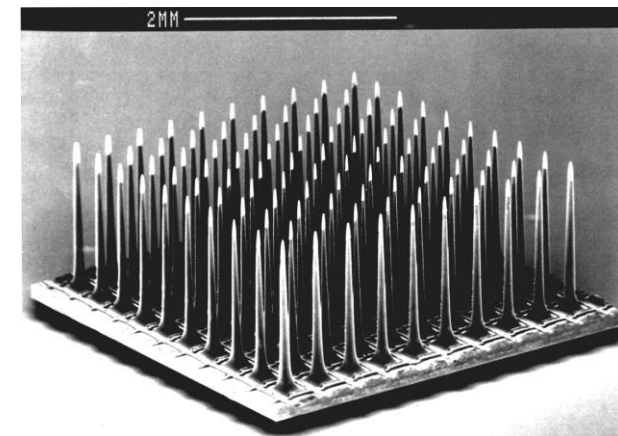






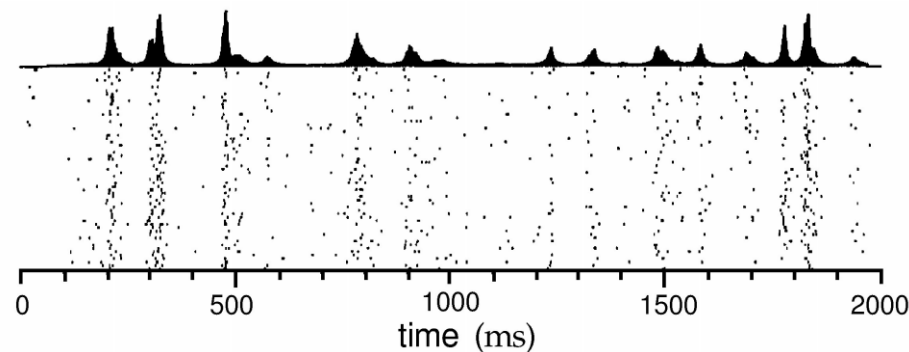
# Agykutatás

- $10^{12}$  (trillion) Neuron
- $10^{15}$  (quadrillion) Szinapszis



- egy sejt másodpercenként 1-500 tüzlés

- Feladat:
  - kapcsolat,
  - funkció feltárás,



- $10 \text{ min} = 64 * 500 * 600 * 1000 \text{ b} \sim 20 \text{ Gb}$

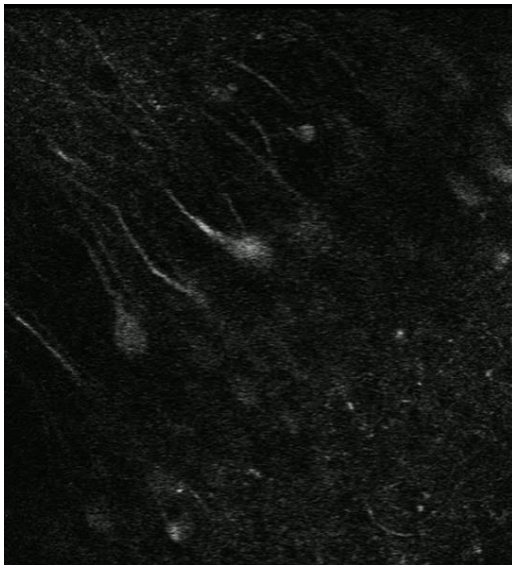


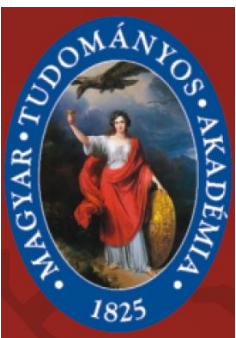
NAP Pattern group

```
10P110100110111
00A011000001011
10T001010111001
01T100101110100
00E111001001101
00R010110000010
10N110100110111
```

# Agykutatás

nyers video



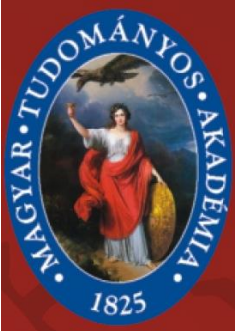


# Számítás a felhőben

- 1 PC 4-8 processor 16-64 Gb memória, <2T tároló
- Felhő lényegében korlátlan, MTA/Wigner
- 1000 Processzor, 8 processzoronként 16-64 Gb Memória 2000 T tároló

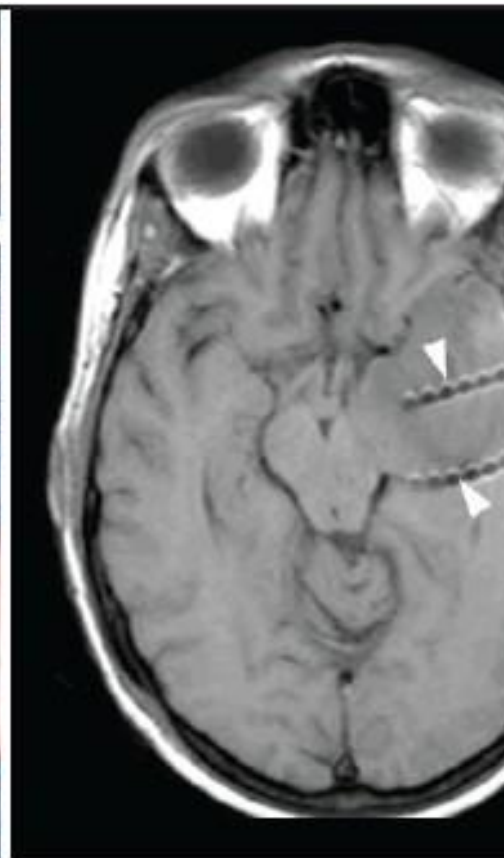
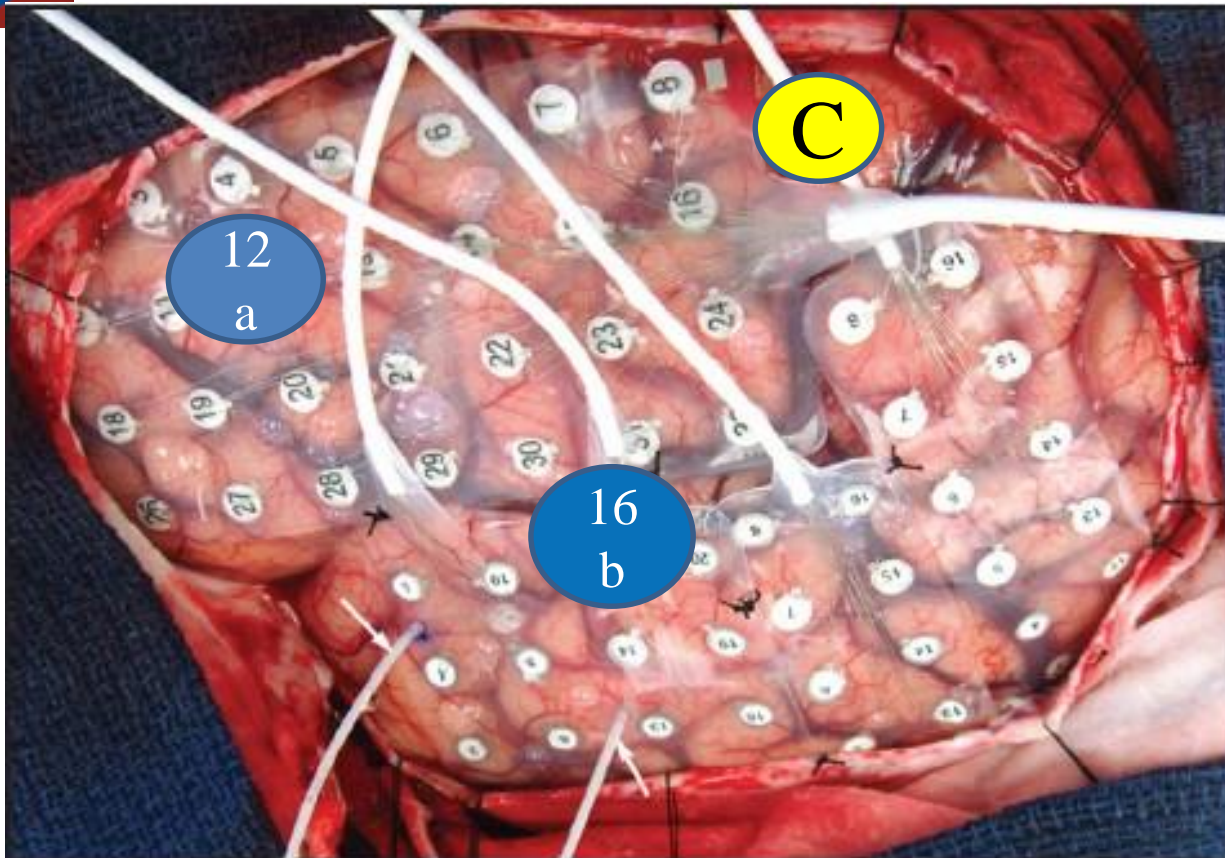




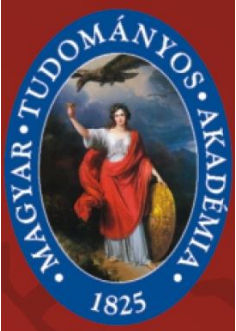


# Még egy példa a felhőben számítás szükségességéről





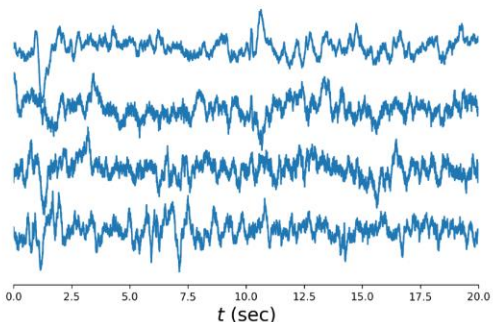
Shah AK, Mittal S. Invasive electroencephalography monitoring: Indications and presurgical planning. Ann Indian Acad Neurol 2014;17, Suppl S1:89-94



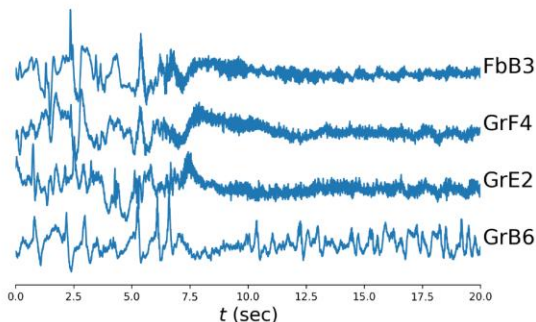
# Epilepszia fókuszának behatárolása



## Asymptomatic



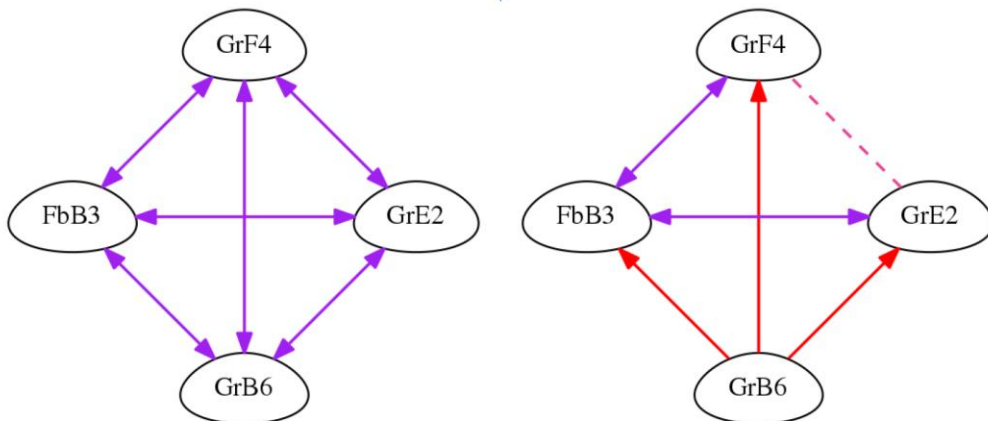
## Epileptic seizure



2 – Data preprocessing  
Band-pass filtering (1-30 Hz)  
Normalization

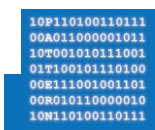
## 3 – Dimension-causality analysis

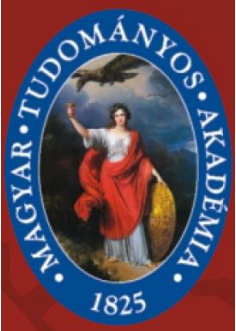
embedding dimension: 5  
embedding delay: 11 step



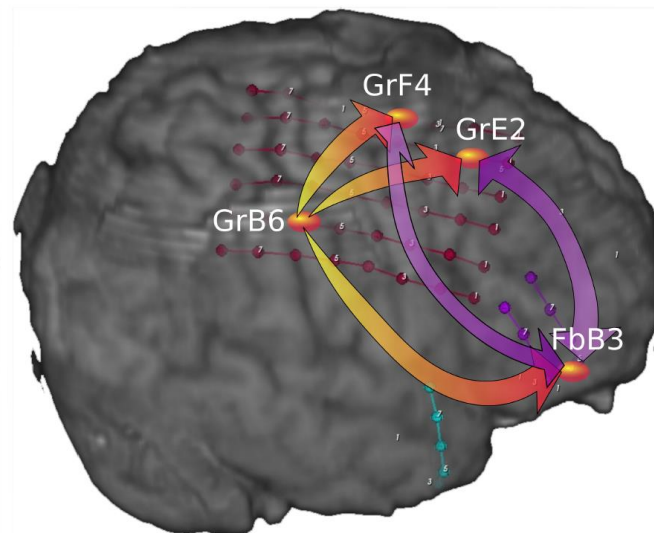
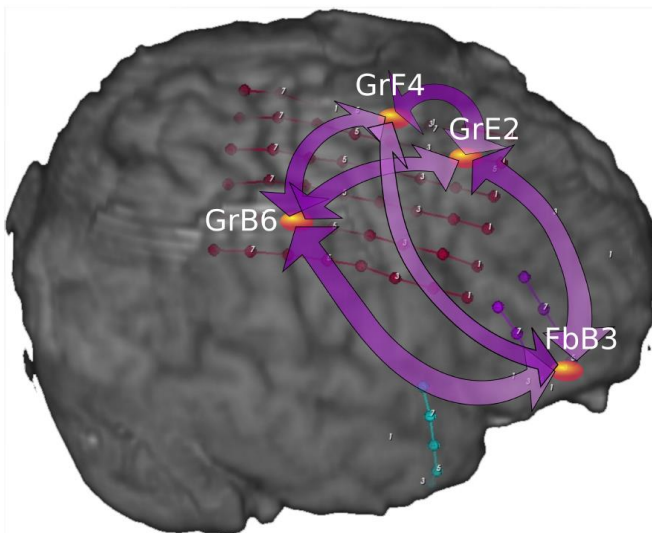
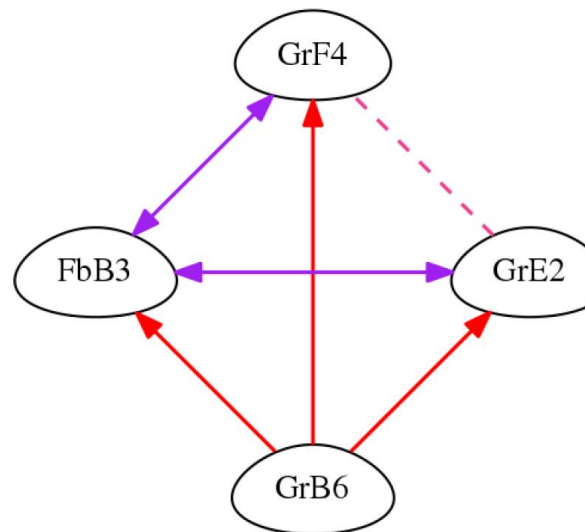
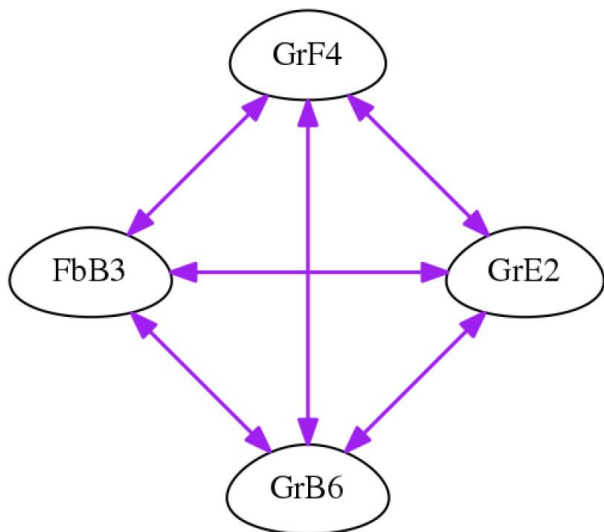
## 4 – Result

Our causality analysis showed that all the 4 area in question were mutually interconnected during normal, interictal activity, but the infero-temporal (GrB6) area became the dominant cause during seizure.

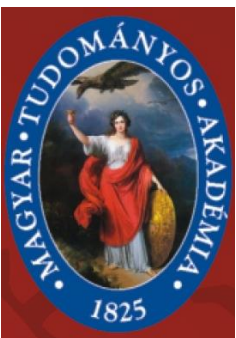




# Epilepszia fókuszának behatárolása







# Felhő

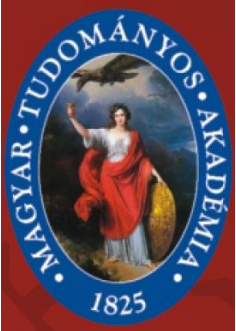
- Egy csatorna-pár 6 sec (két hét kód optimalizálással 2-4 perc helyett)
- 84 csatorna elvezetés (szokásos EEG 64 v 128)
- A párok vizsgálata
- 7056 pár (4096 v 16384)
- 11.76 óra ( 6.8 v 27.3 óra)
- Ez a felhőben 43 sec ( 25 sec v 100 sec)



HUNGARIAN BRAIN RESEARCH PROGRAM  
NEMZETI AGYKUTATÁSI PROGRAM

NAP Pattern group

10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111



Köszönöm a figyelmet



NAP Pattern group

10P110100110111  
00A011000001011  
10T001010111001  
01T100101110100  
00E111001001101  
00R010110000010  
10N110100110111