

Diari de treball – TFG

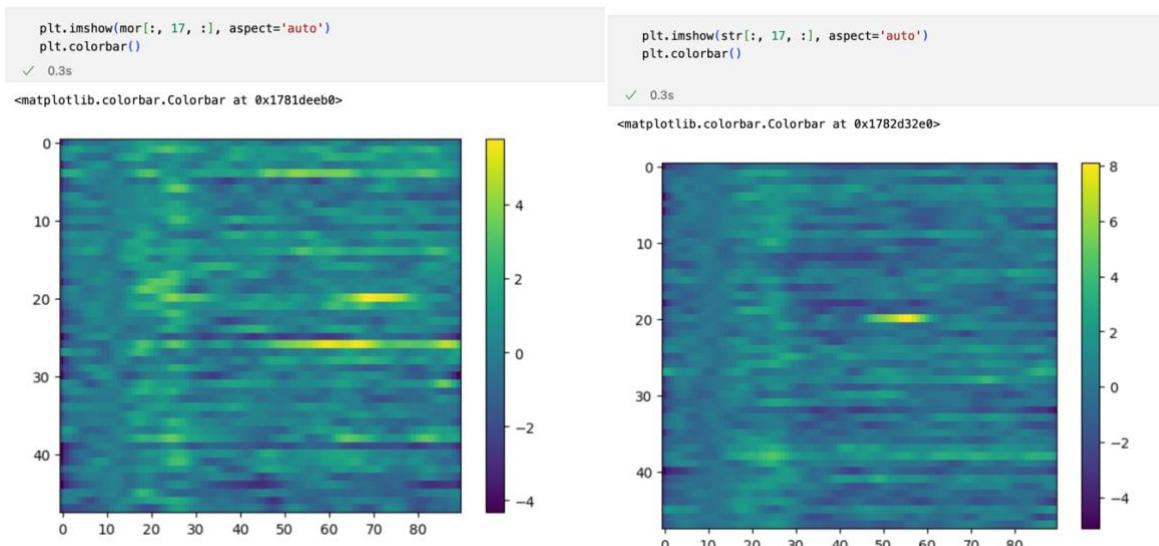
Aquest document ha servit de portapapers durant tot el treball, apunts de reunions, informacions per enviar o consultar, recordatori de feines que he fet, totes aquestes coses conformen aquest document informal on he enganxat totes les idees que tenia i esborrat moltes d'elles. Tot el material és INFORMAL, i ha servit únicament com a suport personal.

Divendres, 14 de Març

He fet una Matriu avg de matrius de correlació de la qual he extret 2 comunitats louvain, d'aquestes, he trobat els canals més importants (18 i 47).

Decidim fer nomes dues comunitats ja que sembla que la majoria de canals es queden en dues comunitats, però farem proves amb la informació dels canals 17 i 21, ja que sempre son els ultims amb diferència de posicionarse.

Vejem visualitzacions dels canals 17 i 48, per a cada trial de mora i stress. És clar que no hi ha pràcticament diferència, i serà molt difícil per a un classificador diferenciar.



Amb aquests he fet una classificació SVM (només amb mean, variance , que ha obtingut molt mal resultat, amb c=1, i per tant, risc de overfitting, només tenim 50% de accuracy en training. Falten moltes dades per a entrenar el model.

Provarem svm one versus all;

Ens trobem, com era de esperar que el model no és capaç d'aprendre, decideix sempre pel grup majoritari. Si entrenem amb 24 syl, 24 stres i 48 mora, per fer mora vs all, igualment no aprèn a decidir, i dona un acc de 34%.

David Martí

Provarem amb finestres temporals, derivades, altres features. Potser discretitzant el valor de power en poques categories o fent allò de la variança amb valors binaris podem passar la seqüència sencera en lloc de les features que estem usant ara.

Dimecres 26 de Març

Crear unes funcions bàsiques de visualització , he fet un arxiu anomenat visualitzacions.ipynb per a carregar les funcions i no haver de escriure cada vegada.

Vejem que el codi de louvain fet er a cada matriu de correlació individualment dona resultats semblants, tot i que no sempre iguals, tot i així no s'explica que el canal 7 o 20 sempre estiguin junts amb la resta

He tractat de modificar el codi de la louvain communities però no trobo l'error tot i veure que actua raro, potser es la funció que vaif haver de canviar per la nx, provaré de fer la separació per comunitats d'una altre manera per a poder avançar.

Canviar trials x subjects al codi. Fet

Labelsss. Fet

Dijous 27 de Març

He fet una funció per a veure diferencies entre canals individuals a visualització

Revisar codi louvain

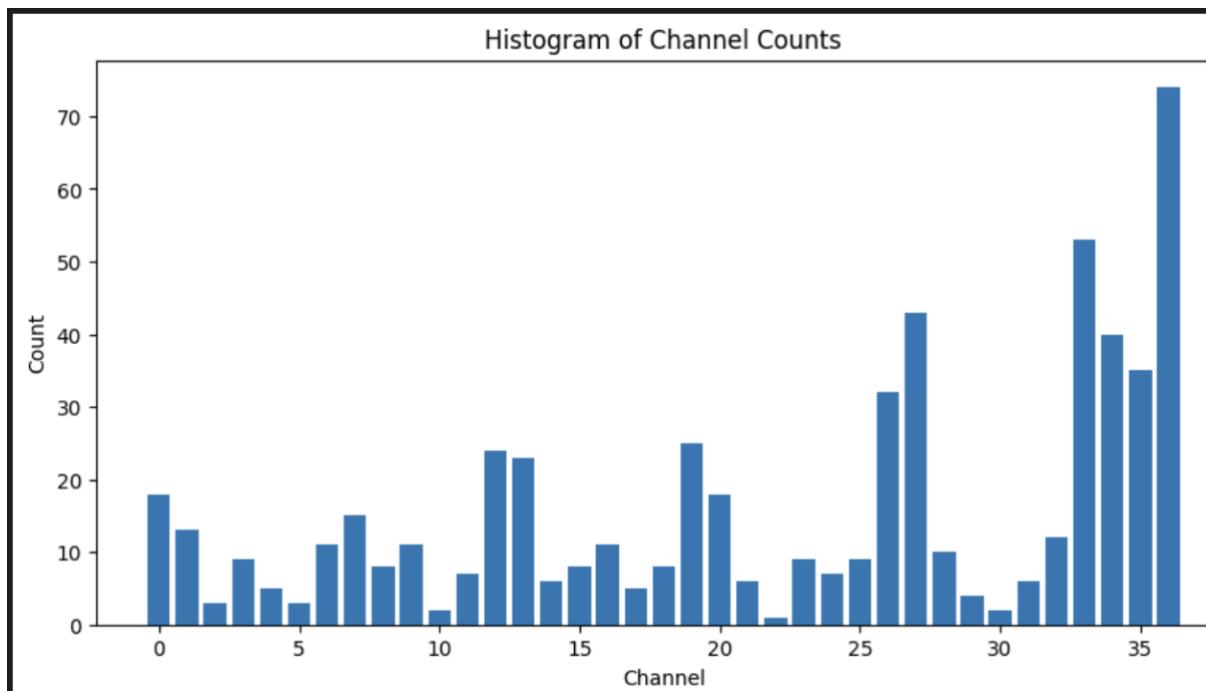
Fer louvain per cada subject i després decidir les comunitats manualment. Comparar amb la mtrr avg. Així es veuen les comunitats de louvain amb el codi que tenim.

```
[{11}, {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 21}, {17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20}, {13, 14}, {21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32}, [{0, 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 16, 17}, {34, 35, 36, 8, 13, 14, 15, 20, 21, 22, 23, 27, 28, 29, 30}, {32, 33, 12, 18, 19, 20, 21, 22, 23, 24, 27, 28, 29, 30}, [{0, 2, 3, 4, 7, 8, 9, 17}, {1, 33, 5, 6, 10, 11, 12, 18, 19, 25, 26}, {13, 14, 15, 16, 20, 21, 22, 23, 24, 27, 28, 29, 30}, [{1, 5, 6, 10, 11, 12, 16, 17, 18, 19, 24, 25, 26, 31, 32, 33, 35}, {0, 2, 3, 4, 7, 8, 9, 13, 14, 15, 20, 21, 22, 23, 27, 28}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21}, {32, 33, 34, 35, 36, 22, 23, 24, 25, 26, 27, 28}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 25}, {32, 33, 34, 35, 36, 15, 20, 21, 22, 23, 24, 26, 27}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 27}, {10, 15, 16, 17, 20, 21, 22, 28}, {32, 33, 34, 35, 36, 12, 18, 19, 23, 24, 25}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17}, {12, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 25}, {20}, {32, 33, 34, 35, 36, 21, 22, 23, 24, 26, 27, 28}, None], [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 21}, {18, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 21, 27, 28, 33, 36}, {11, 13, 14, 16, 17, 18, 19, 22, 23, 24, 25, 26, 29, 30}, [{1, 5, 6, 10, 11, 12, 17, 18, 19, 23, 24, 25, 26, 30, 31, 32, 33, 34, 36}, {0, 2, 3, 4, 7, 8, 9, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}], [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 20, 21, 23, 27}, {32, 33, 34, 35, 36, 18, 19, 24, 25, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16}, {20}, {17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20}, {16, 24}, {32, 33, 34, 35, 36, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 25}, {13, 6}, {18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{31}, {1, 4, 5, 6, 10, 11, 12, 16, 17, 18, 19, 24, 25, 26, 32, 33, 36}, {0, 2, 3, 7, 8, 9, 13, 14, 15, 20, 21, 22, 23, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17}, {24, 21}, {18, 19, 20, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20}, {33, 18, 19, 25, 26}, {32, 34, 35, 36, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 19, 20, 25, 26}, {24}, {1, 15, 16, 17, 21, 22, 23, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37}, [{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 18, 19}, {11, 15, 16, 17, 21, 22, 23, 24, 25, 29, 30, 31}, {32, 33, 34, 35, 36, 37}], [1] ✓ 0.3s]
```

He aplicat PCA de dues maneres, tant globalment:

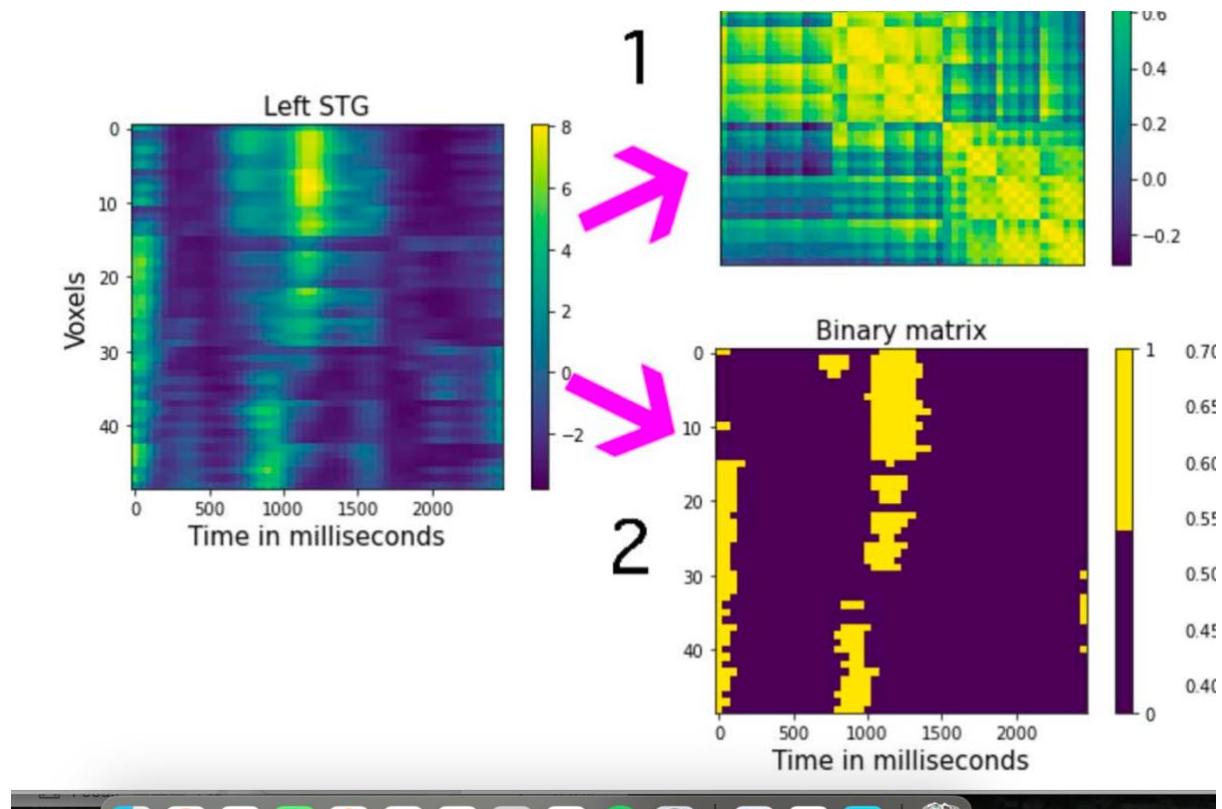
```
• for ch in top_global:  
    print(channel_info[central_chans[ch]])  
  
[1] ✓ 0.3s  
  
• Canals globals: [22, 34, 26, 36]  
(37, b'CP1', -0.0919468, -0.0844667)  
(53, b'P03', -0.11701769, -0.27925784)  
(41, b'CP6', 0.28041118, -0.09507808)  
(55, b'P04', 0.11701769, -0.28009445)
```

Com de forma individual i sumant els canals més importants

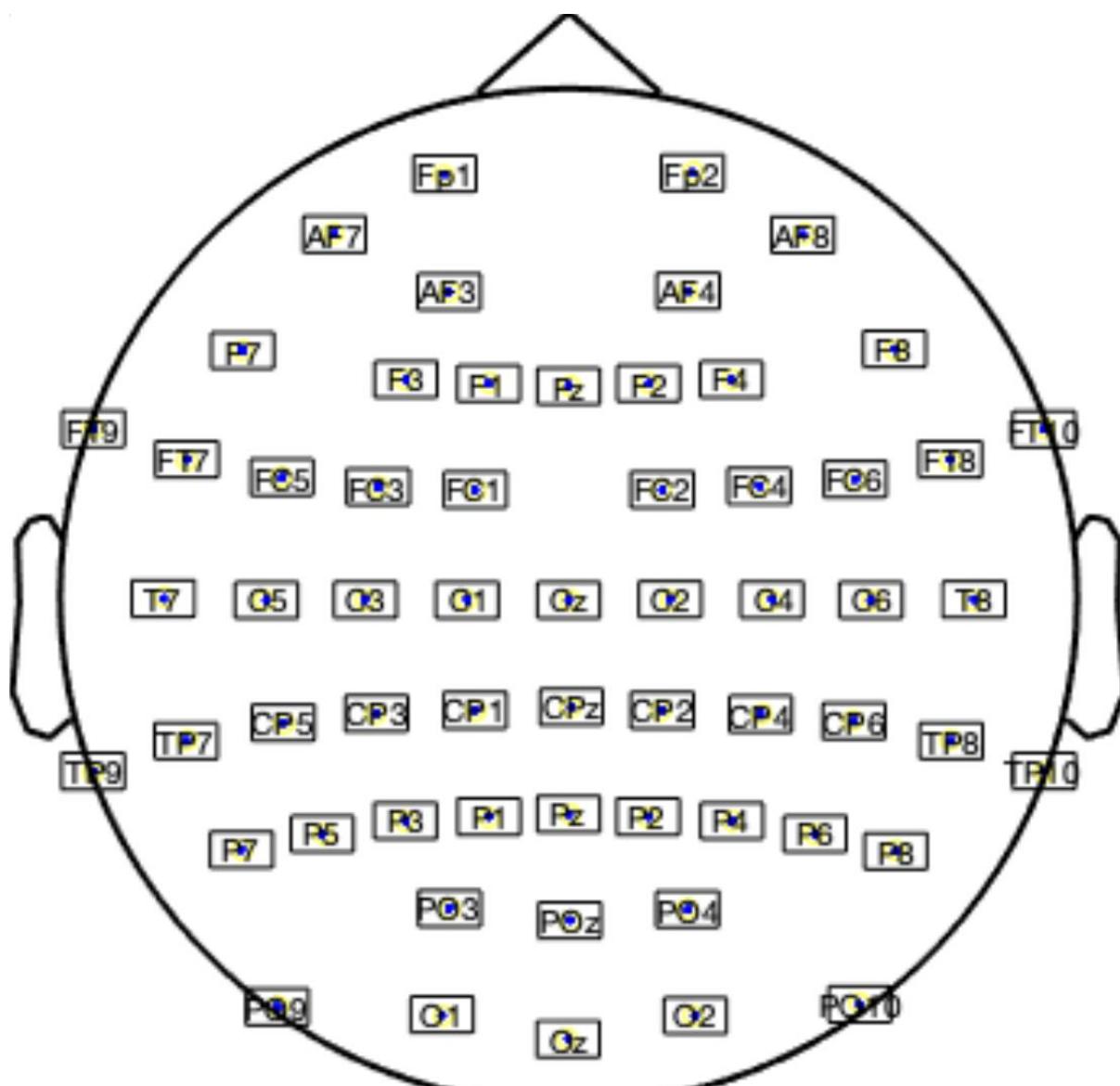


Comentaris reunions

No cal fer scalar, cada valor que tinc es com ha variat la potència respecte el valor dels baseline, trobem la normalització en dades log ja que tenim power i no amplitud.



Caldri trobar un frontal un drte i un esquerra significatius, si es així, te sentit i podem fer una classificació amb aquests senyals.



A partir de lots arribar a la conclusió de que els frontals es comporten iguals els esquerra (imoars) també i dreta (pars) també!!

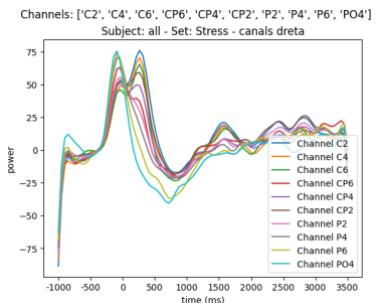
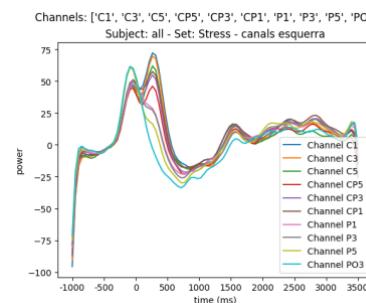
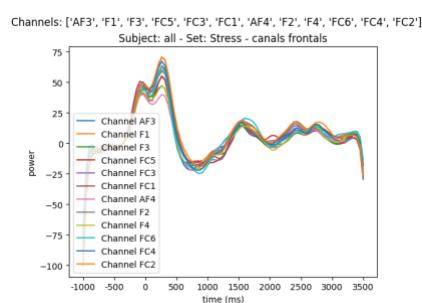
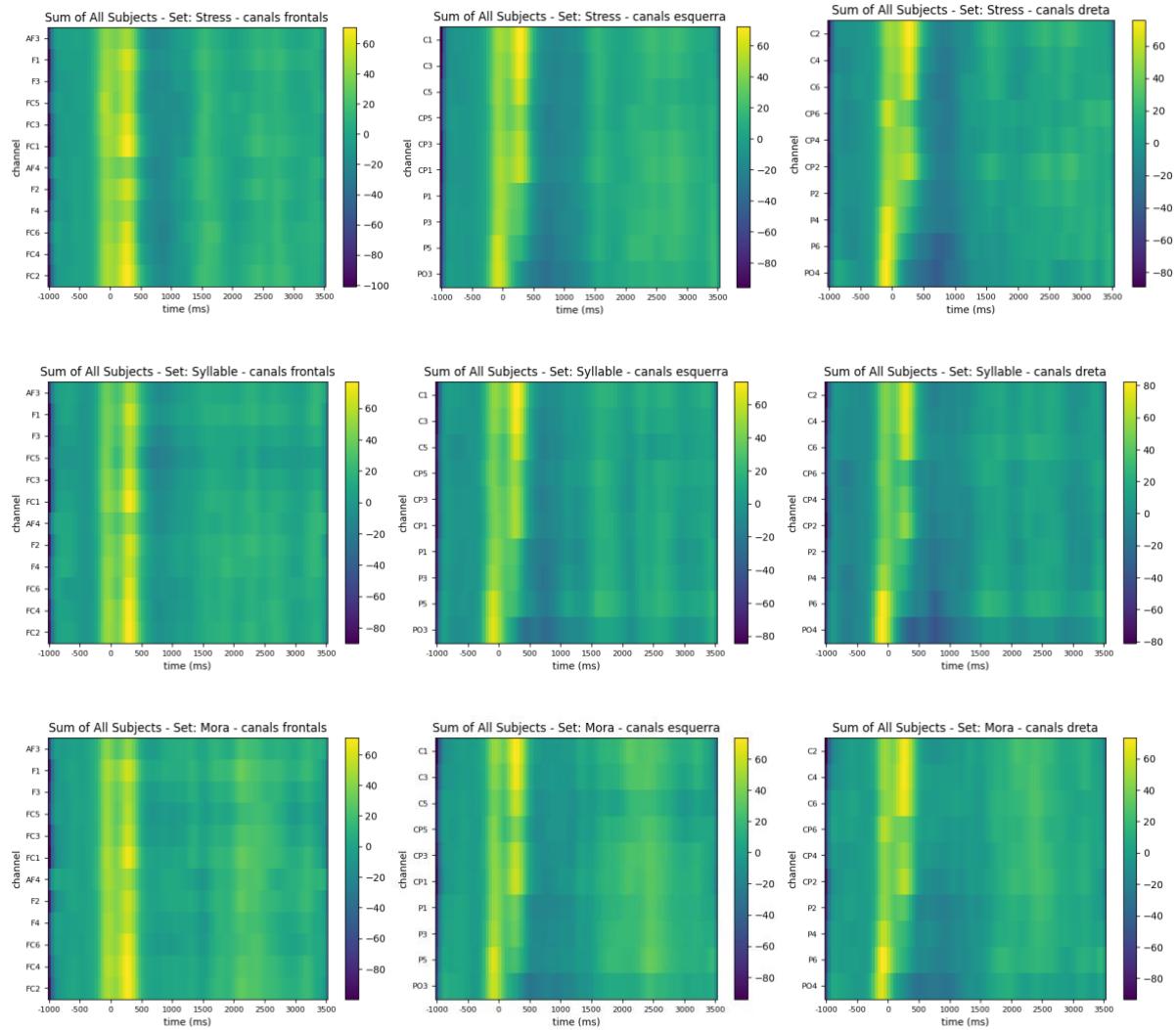
Amb combinació del pca podem ja triar els canals

Ticks del temps a totes les visualitzacions

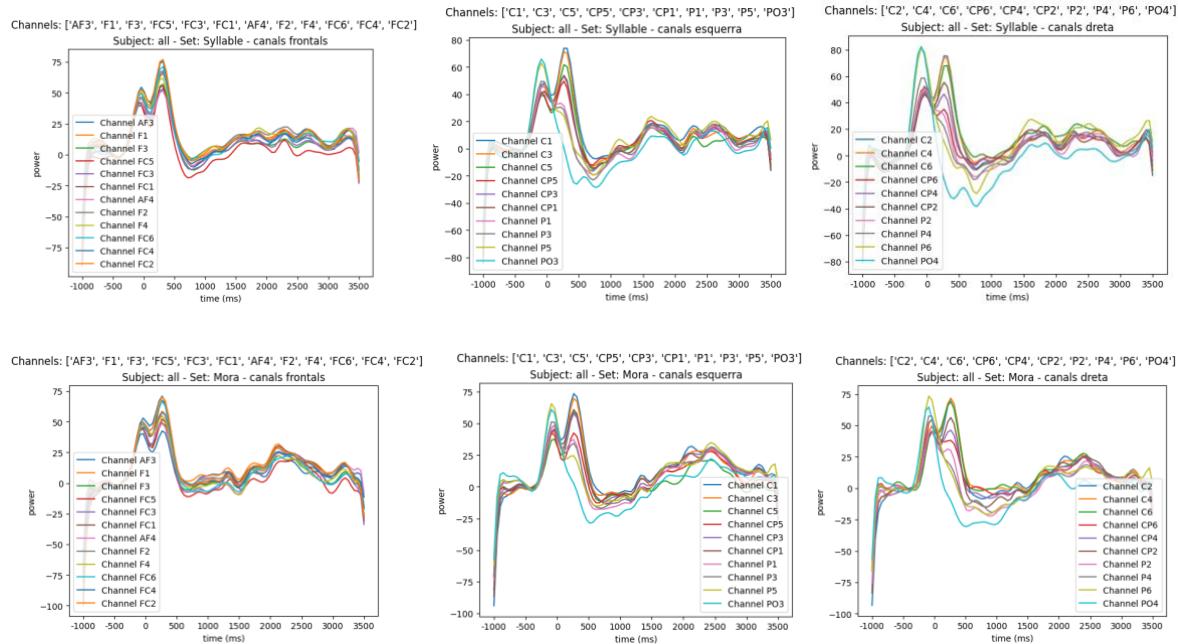
Dissabte 5 d'abril del 25

Avui he ajustat els tick de tot el que he vist, he fet diferents visualitzacions que ara enganxaré i he creat un diccionari per a guardar tots els valors dels noms dels canals i els indexs. Udsant el mapa de canals al cervell, he fet visualitzacions de la part frontal, esquerra i dreta i aquests són els resultats.

David Martí



David Martí

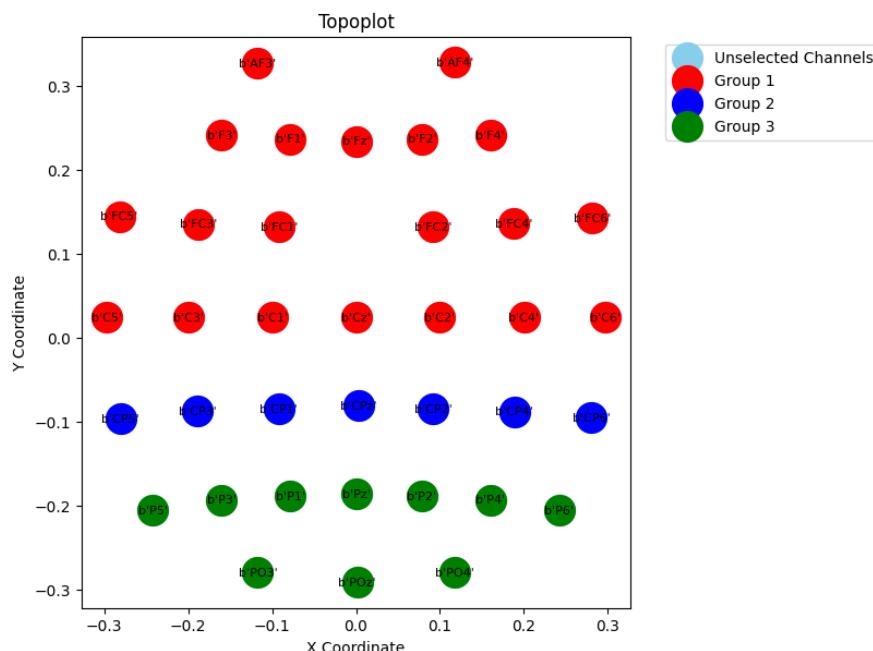


Divendres 11 d'Abril

He fet la funció per a poder fer el topoplott

Dimarts 15 d'Abril

He fet les comunitats amb una matriu que guarda les vegades que louvain guarda cada parella de canals junts, normalitzada. Creant un graf i eliminant les arestes més baixes fins a tenir 3 comunitats.



Dimecres, 16 d'abril

David Martí

Avui he descarregat les dades trial, i les he emmagatzemat en la carpeta trials. Es defineix per a poder usarles, els següents passos:

Slicing i avg en les dimensions

Slicing en els moments de temps

Les dimensions son molt grans pel que un cop fet s'esborraran les dades de la carpeta per a no llegir tota l'estona totes les dades.

Sí, en principi jo em quedaria amb la banda 5-8 Hz, i el temps de 500 a 3500 ms. Per això hauràs de fer slicing i average en la dimensió de la freqüència, i només slicing en la de temps.

Per a cada subject i condició tindràs al voltant de 37-40 trials.

Dijous 17 d'abril

Avui he baixat totes les dades trial, les he carregat a visualitzacions, i ara es poden accedir desde qualsevol fitxer que imorti visualitzacions. He canviat els noms per tal que els avg es diguin avg i a esp_mor, etc guardar les dades completes de trial.

També he fet a la matriu de correlació les dades trial, la comunitat louvain i he pintat les comunitats. Però no es consegueixen veure els 3 grups diferenciats. Cal pensar que fem o quina conclusió li dono.

Topoplot amb els canals activats : codi encara. Fet

Matriu amb veinalitat de canals i després

Juntar per valors mes alts fins que quedin 3 grups

Explicar que es la resolució, que es buscava i que hem trobat.

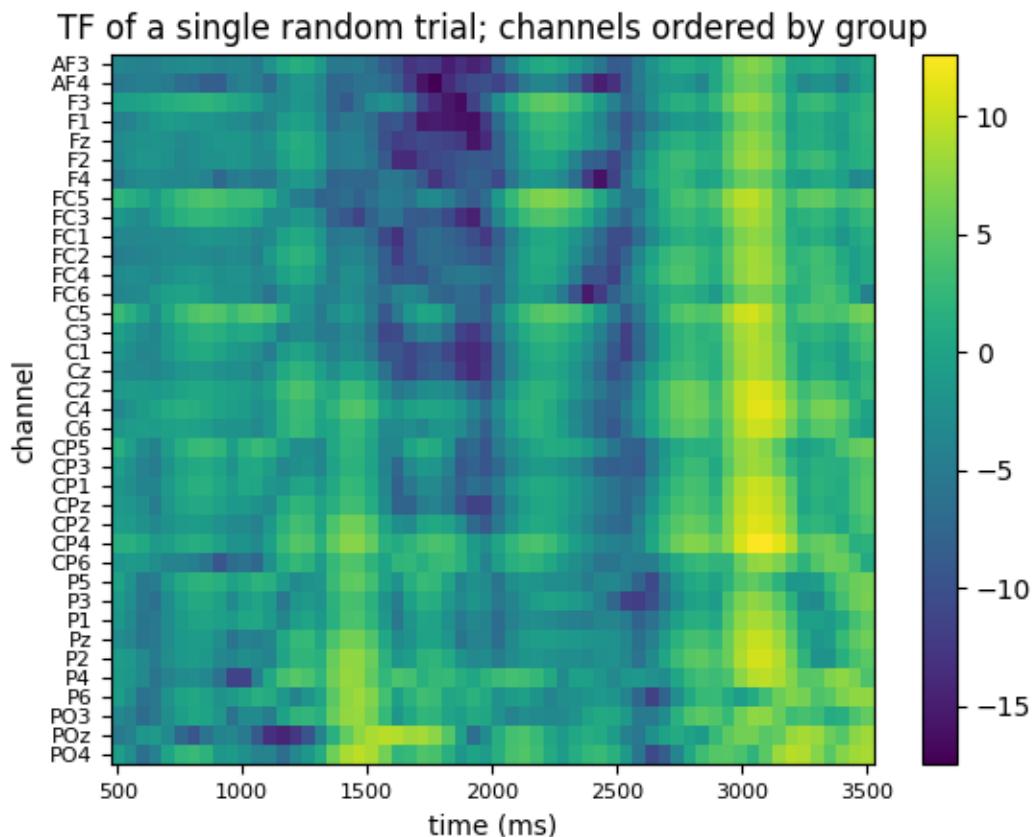
Binaritzacio amb std dev com explica al paper → classificar

Utilitzar histogrames i tots els gràfics que poguem per fer la classificació.

25 d'Abril

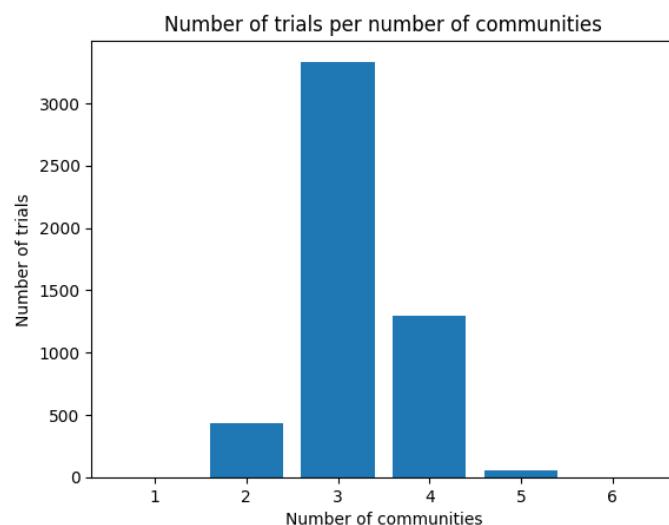
ordre =

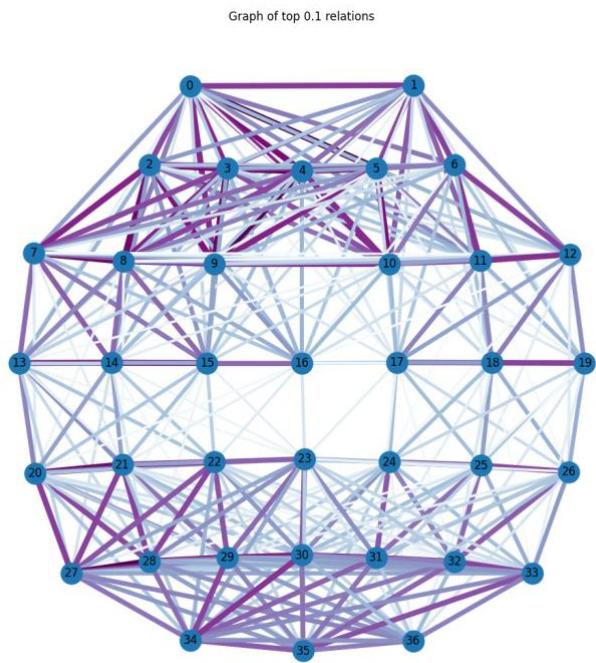
[1,2,3,4,5,7,8,9,10,13,19,32,34,6,11,12,17,18,24,25,26,30,31,14,15,16,20,21,22,23,27,28,29,33,35,36]



Dijous 1 de Maig

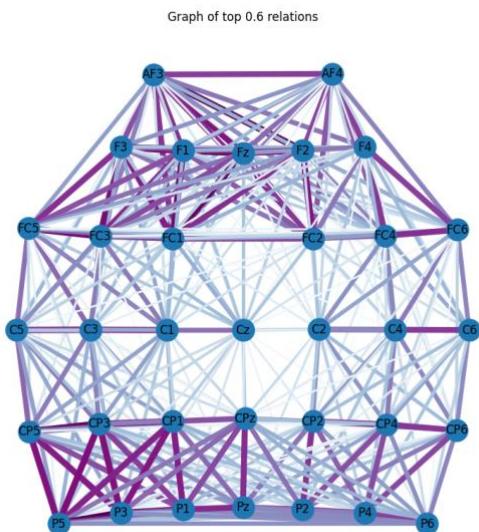
Avui he revisat tot el codi que tinc fet per a la detecció de comunitats. Res sembla estar mal fet. Adjunto histograma del nombre de comunitats per a tots els trials i graf de veinalitat, que ens fa a entendre que hi ha dos grups clars que es relacionen mes que la resta.





A partir d'aquí deixo d'apuntar la data:

Avui he canviat i retocat Louvain al graf anterior, ara utilitza una resolució diferent i permet diferenciar els 3 grups de forma més clara. A més he esborrat els 3 canals de la funció de visualització.



Ara tenim més solidesa per a la classificació del model.

Comencem a treballar en el desenvolupament de les features per al model

Tres grups de canals segons la classificació (esquerra, dreta, frontals)
Fer la mitjana dels canals per grups

Binaritzar la senyal, i trobar estadístiques que puguin ser útils:

- La derivada
- Finestres temporals
- ..
- ..
- ..

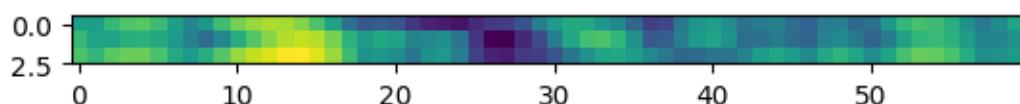
Definir un model de classificació simple.

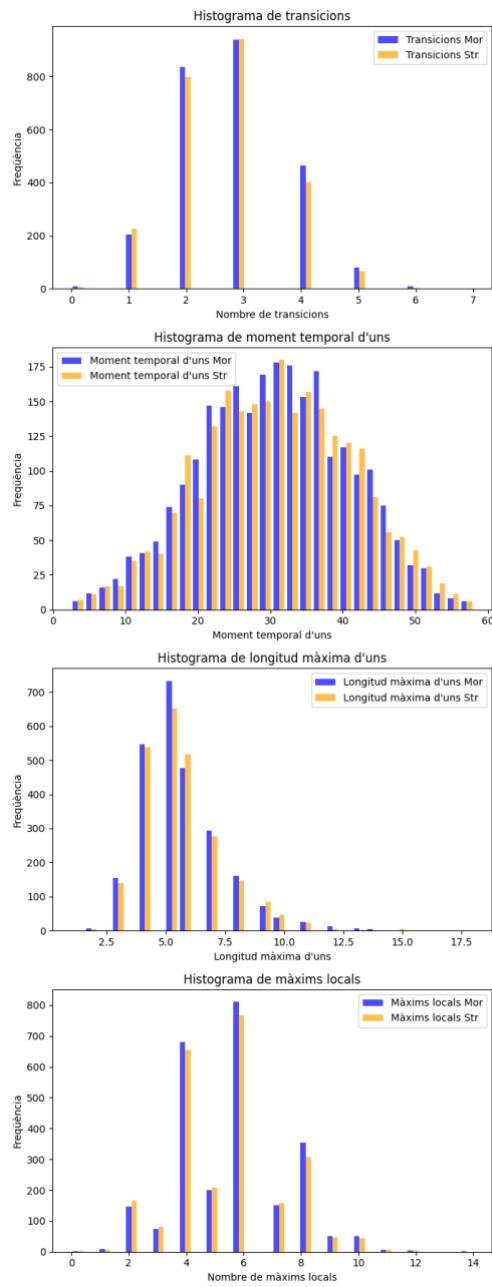
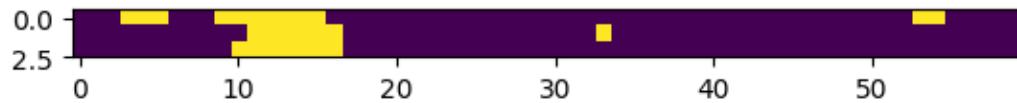
Possibles features:

Proporció d'estats alts (uns): Fracció de punts de temps on el senyal és 1 (per sobre del llindar). Aquesta és una estadística bàsica (equivalent a la mitjana del binaritzat) que indica quant temps el senyal està actiu. Exemple sintètic: per la seqüència [1,0,1,1,0] hi ha 3 valors 1 sobre 5 punts, per tant la proporció és 0.6 (60%).

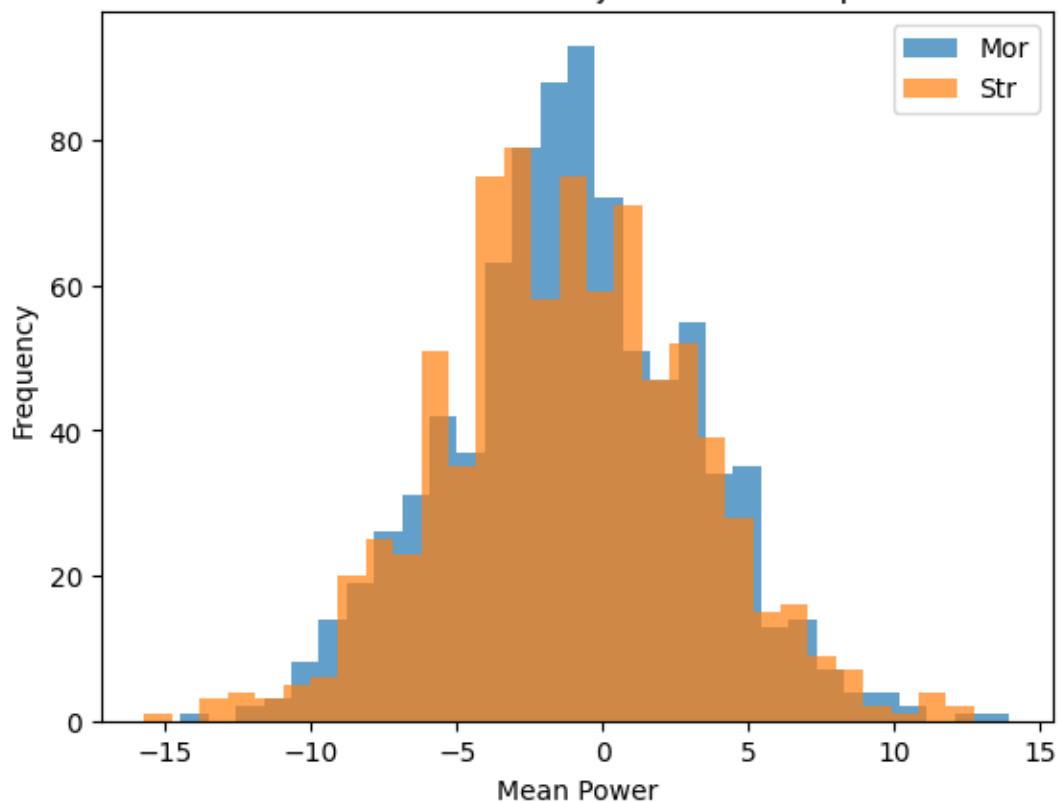
- Nombre de transicions 0↔1: Comptar quantes vegades el valor canvia d'0 a 1 o de 1 a 0 al llarg de la sèrie. Un nombre alt de transicions indica un senyal que alterna sovint, mentre que un valor baix indica períodes més llargs consecutius d'estat constant. Exemple: en [1,0,1,1,0], les transicions són 1→0 (índex 1→2), 0→1 (2→3) i 1→0 (4→5), un total de 3 transicions.
- Nombre de ratxes (runs) d'1 (i de 0): Nombre de blocs consecutius de valors 1 (i de valors 0). Cada ratxa comença quan el valor canvia (veure transicions). Proporciona informació de quantes vegades el senyal passa repetidament a l'estat 1 o 0. Exemple: la seqüència [1,1,0,1,0,0] té 2 ratxes de 1s (11 i 1) i 2 ratxes de 0s (0 i 00).
- Durada mitjana de ratxes d'1 i de 0: Longitud mitjana dels blocs consecutius d'uns (i de zeros). Mesura la persistència típica de cada estat. Exemple: per [1,1,0,1,0], els blocs de 1s són de longituds 2 i 1 (mitjana 1.5) i els blocs de 0s de longituds 1 i 1 (mitjana 1.0).
- Longitud màxima de ratxa d'1 (i de 0): Longitud del bloc més llarg de 1s (i de 0s). Captura la ratxa prolongada més gran, que pot diferenciar senyals amb una gran oscil·lació d'aquells amb breus interrupcions. Exemple: [1,0,0,1,1,1,0] té una ratxa màxima de 1s de 3 (els tres últims 1s consecutius) i una de 0s de 2.
- Moments temporals dels 1s: Primer moment (posició mitjana) i, si cal, segon moment (dispersió) de les posicions on apareix el valor 1. Això indica cap a quina part de la finestra de temps tendeix a concentrar-se l'activitat. Exemple: en [0,1,0,0,1] els 1 apareixen a les posicions 2 i 5 (sobre 5 punts), amb posició mitjana $(2+5)/2 = 3.5$, suggerint que els punts actius es concentren cap a la meitat-final de la finestra. Una variança alta implicaria que estan més dispersos.

Cañí expñicar la binarització:

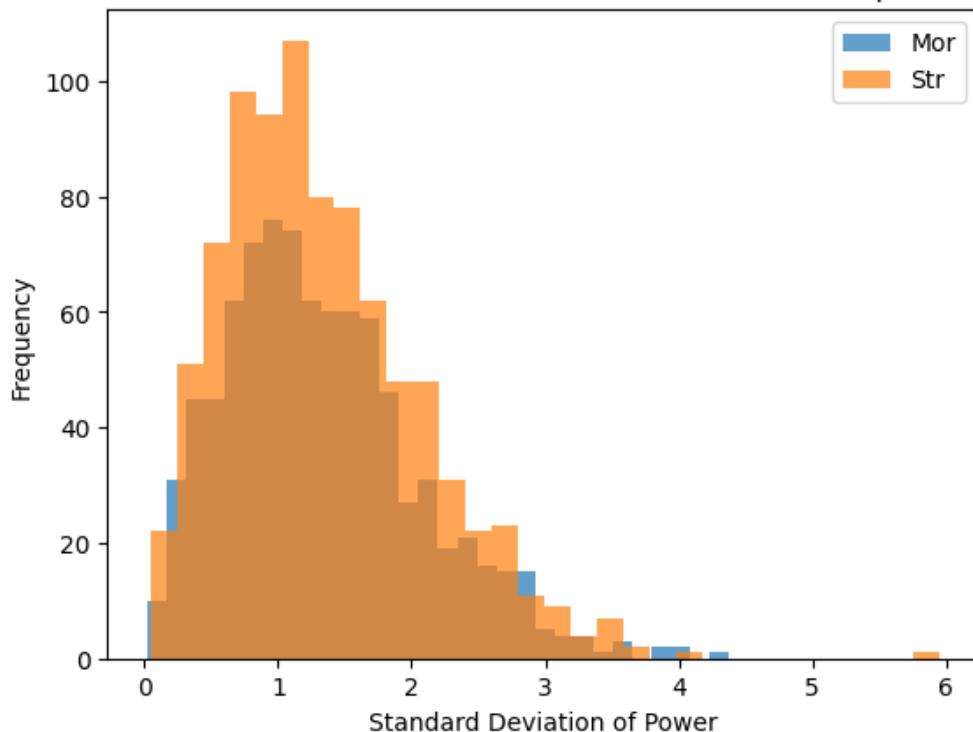




Distribució de la Potència Mitjana en els Grups Mor i Str



Distribució de la Desviació Estàndard de la Potència en els Grups Mor i Str (ESP)

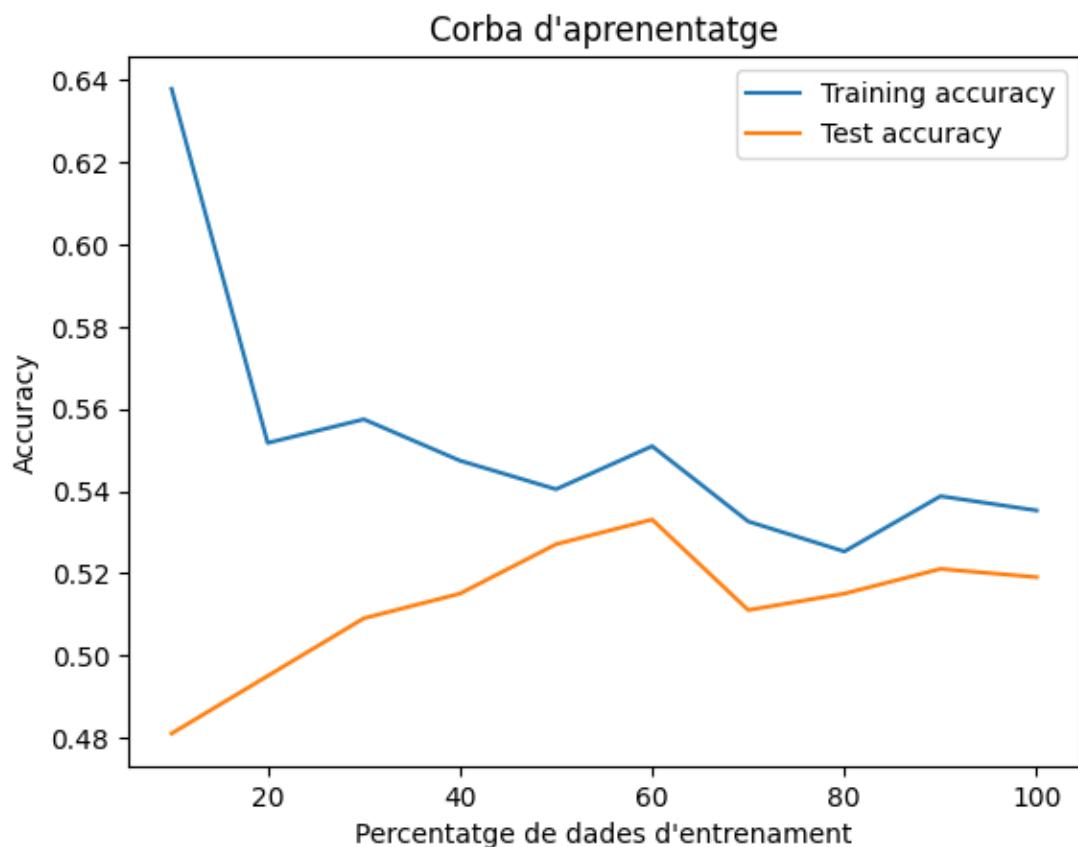


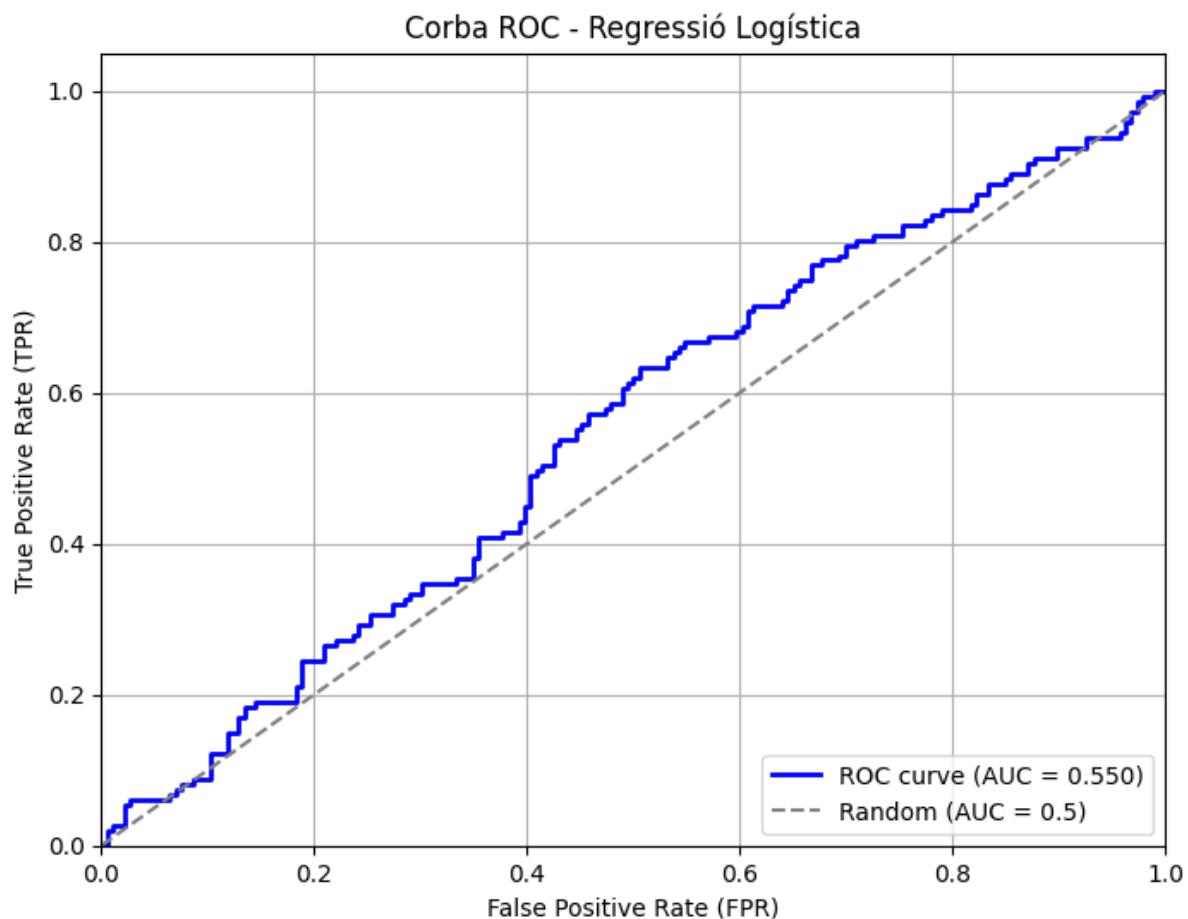
Vegem que cap de les features es realment informativa, tot i així prepararem un model:

Model de classificació:

- Perquè log regression?: molta variabilitat, ... soroll a les dades, ...

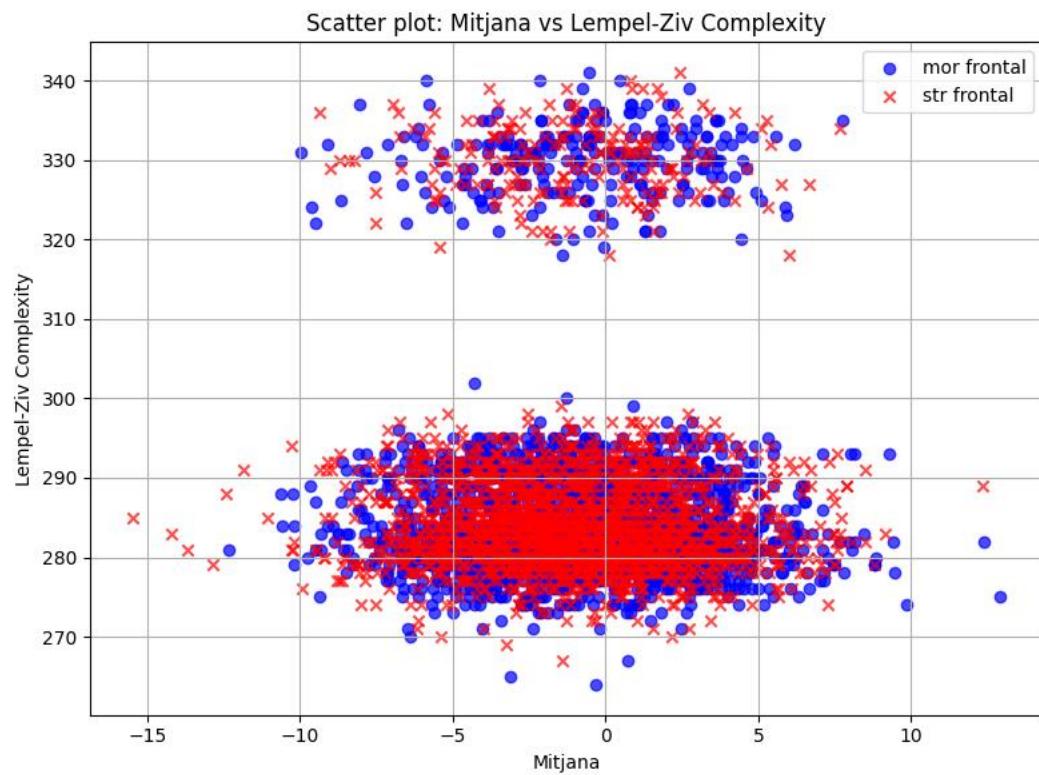
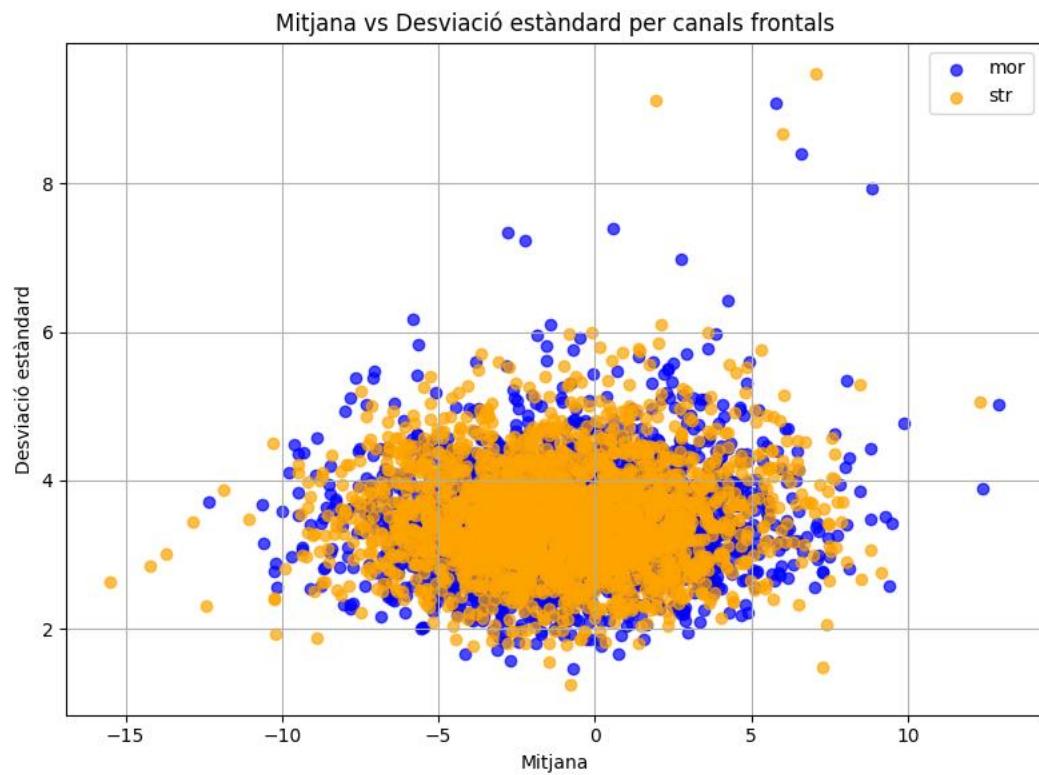
** explicat al informe directament.





```
Accuracy: 0.5135135135135135
F1 Score: 0.5149700598802396
Confusion Matrix:
 [[85 92]
 [70 86]]
```

Analisi de features: Mitjana: Recompte de quantes vegades la mitjana de les dades



Les features no expliquen res. Potser PLV? Seria fer trampa?? Ja que es calcula amb les dades de el “so”.

Calcular la complexitat enviant les dades en columna (1 sola srting de 180 on l'ordre es 111222333444...

David Martí

Bibliografia

1. Feature extraction for EEG classification using time-domain parameters

Cita:

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4), 1084–1093.

URL:

<https://doi.org/10.1016/j.eswa.2006.02.005>

Ús en el teu cas:

Inspiració per utilitzar característiques temporals simples com la mitjana, desviació estàndard, màxim i mínim sobre finestres de senyal. Aquest enfocament encaixa amb la teva idea de mantenir el model senzill però informatiu.

2. Hjorth Parameters as Temporal Features

Cita:

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3), 306–310.

URL (Wikipedia, resum actualitzat):

https://en.wikipedia.org/wiki/Hjorth_parameters

Ús en el teu cas:

Es va fer servir per suggerir una forma de caracteritzar les transicions i la “activitat temporal” del senyal després de binaritzar o abans, com a alternatives a mitjanes simples. Ideal per model senzill amb EEG.

3. Scikit-learn for simple models like Logistic Regression

Cita:

Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

URL:

<https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>

Ús en el teu cas:

Base per recomanar l’ús de regressió logística i SVM com models inicials interpretables per a classificació multinomial amb poques features.

4. On Binarization and Derivatives in Signal Processing

Cita:

David Martí

No hi ha una font específica que s'hagi citat literalment aquí, però el concepte ve de pràctiques comunes en senyals binaritzats per detectar punts de canvi o patrons. L'ús de la derivada com a “feature” abans de binaritzar ve inspirat per tractament de senyals en estudis com:

Cita recomanada:

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, 250, 34–46.

URL:

<https://doi.org/10.1016/j.jneumeth.2014.08.002>

Ús en el teu cas:

Per suggerir que derivar abans de binaritzar pot destacar punts de canvi o patrons transitoris en el senyal EEG.