

Análisis topológico en un estudio de neurociencia del comportamiento

Carles Casacuberta, Ignasi Cos, Aina Ferrà

15 de noviembre de 2022



UNIVERSITAT DE
BARCELONA

Motivación

La motivación y la obtención de recompensas son dos de los principales motores del comportamiento humano.

Objetivo principal

Analizar cómo gestiona la motivación el cerebro.

- Se diseña un experimento para estudiar la motivación modulada por la presión social.
- Se ve que existe una red cerebral que interpreta la presión social como motivación, afectando al comportamiento aun en la ausencia de recompensa.
- **Se quiere entender las diferencias entre los diferentes estados motivacionales.**

Motivación

La motivación y la obtención de recompensas son dos de los principales motores del comportamiento humano.

Objetivo principal

Analizar cómo gestiona la motivación el cerebro.

- Se diseña un experimento para estudiar la motivación modulada por la presión social.
- Se ve que existe una red cerebral que interpreta la presión social como motivación, afectando al comportamiento aun en la ausencia de recompensa.
- **Se quiere entender las diferencias entre los diferentes estados motivacionales.**

Importante

Usar técnicas **explicables** que aporten información adicional.

Estructura de la presentación

- 1 **Análisis topológico de datos**
- 2 Clasificador topológico
- 3 Detalles del estudio
- 4 Resultados
- 5 Conclusiones

Análisis topológico de datos

El **análisis topológico de datos** (TDA) es una rama relativamente nueva de la ciencia de datos que aplica algoritmos de la topología algebraica para encontrar características de un conjunto de datos —principalmente, busca describir **la forma** de una nube de puntos mediante invariantes topológicos.

Tenemos dos maneras principales de trabajar con TDA:

- Métodos cualitativos (**Mapper**): nos dan información de los datos pero sin necesariamente aportar datos numéricos.
- Métodos cuantitativos (**homología persistente**): herramientas muy extendidas que incluso se usan como apoyo en algoritmos de aprendizaje automático.

Análisis topológico de datos

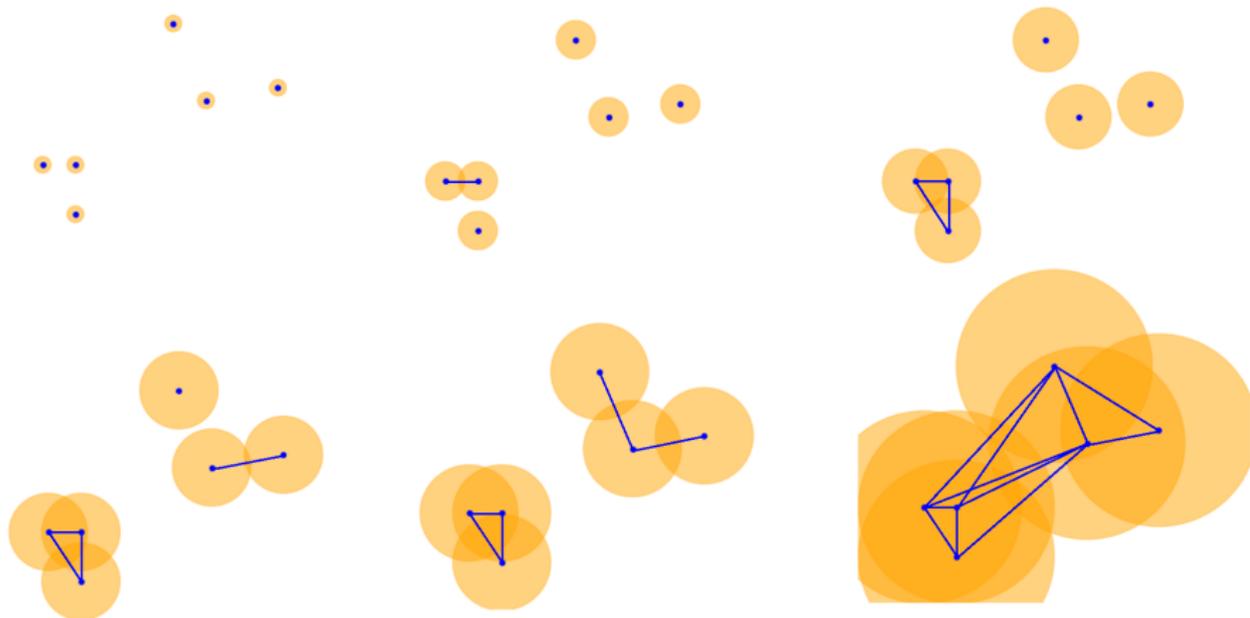
El **análisis topológico de datos** (TDA) es una rama relativamente nueva de la ciencia de datos que aplica algoritmos de la topología algebraica para encontrar características de un conjunto de datos —principalmente, busca describir **la forma** de una nube de puntos mediante invariantes topológicos.

Tenemos dos maneras principales de trabajar con TDA:

- Métodos cualitativos (**Mapper**): nos dan información de los datos pero sin necesariamente aportar datos numéricos.
- Métodos cuantitativos (**homología persistente**): herramientas muy extendidas que incluso se usan como apoyo en algoritmos de aprendizaje automático.

Análisis topológico de datos

Empezamos calculando una filtración sobre nuestra nube de datos. En este caso, contaremos **componentes conexas**.



Análisis topológico de datos

Diferentes maneras de resumir esta información:

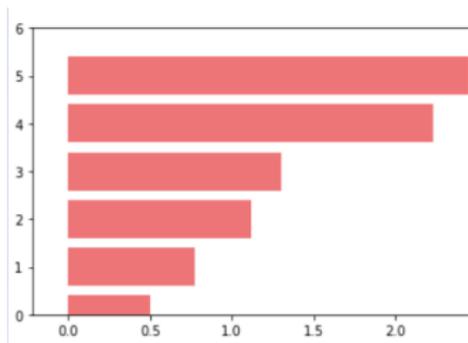


Diagrama de barras

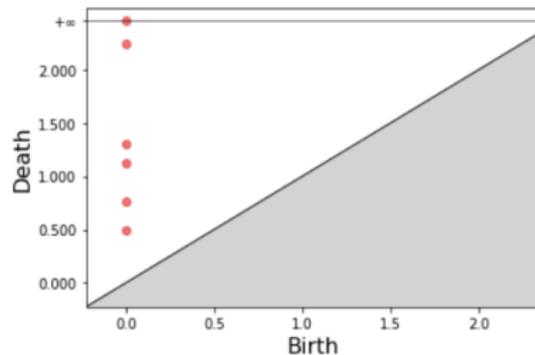


Diagrama de persistencia

Análisis topológico de datos

Diferentes maneras de resumir esta información:

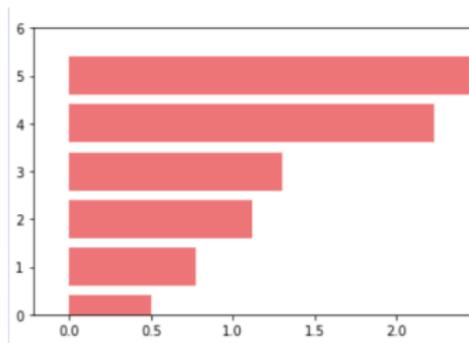


Diagrama de barras

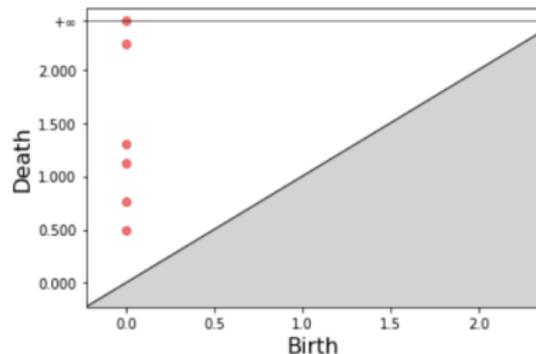
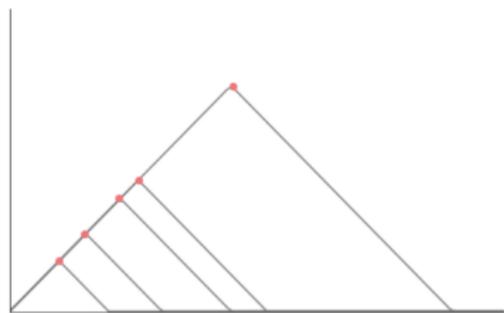


Diagrama de persistencia

Las formas **vectorizadas** son más útiles para trabajar con aprendizaje automático.

Análisis topológico de datos



Perfil de persistencia



Silueta

Funciones triangulares:

$$\Lambda_{(b,d)}(t) = \max\{0, \min\{t - b, d - t\}\}$$

Sucesión de perfiles:

$$\lambda_k(t) = k \max\{\Lambda_{(b,d)}(t)\}$$

$$\phi_w(t) = \frac{\sum_{i=1}^m w_i \Lambda_{(b_i, d_i)}(t)}{\sum_{i=1}^m w_i}$$

Estructura de la presentación

- 1 Análisis topológico de datos
- 2 **Clasificador topológico**
- 3 Detalles del estudio
- 4 Resultados
- 5 Conclusiones

Clasificador topológico

Trataremos con un problema de **clasificación**, es decir, nuestro objetivo es asignar etiquetas a los datos.

En nuestro experimento

Dados unos datos tenemos que asignar a qué estado motivacional corresponden.

Para poder comprobar la eficacia de nuestro clasificador, partiremos nuestro conjunto de datos en:

- 80% entrenamiento, datos de los cuales sabemos las etiquetas.
- 20% test, datos de los que suponemos que no conocemos las etiquetas y debemos etiquetar nosotros.

Como medida de eficacia calcularemos la **precisión**: porcentaje de datos del conjunto de test correctamente clasificados.

Clasificador topológico.

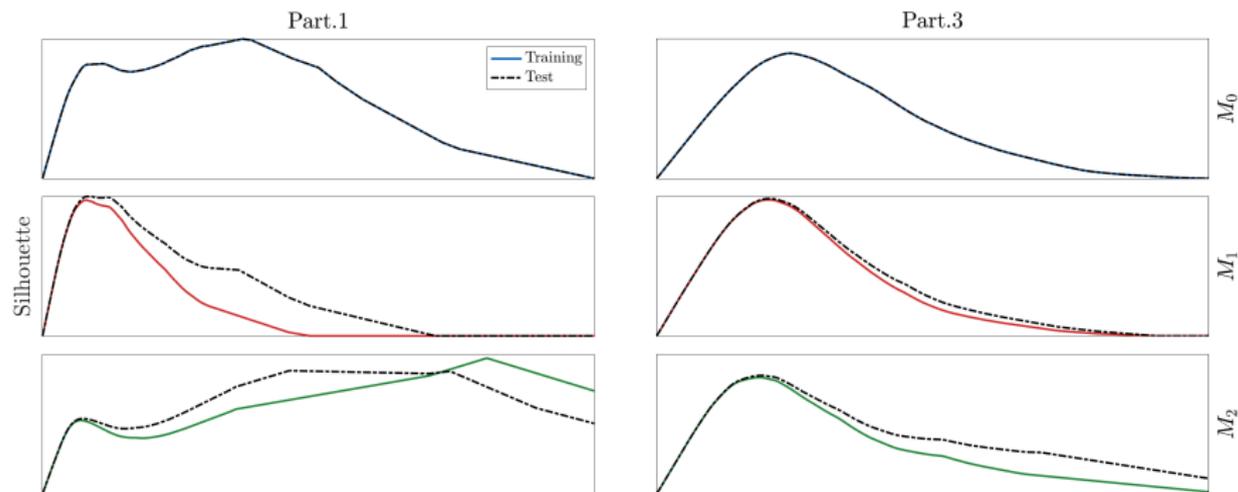
Proponemos el siguiente algoritmo.

Clasificador persistente

- 1 Separamos el conjunto de entrenamiento según las clases.
- 2 Para cada clase, calculamos su silueta en dimensión homológica 0.
- 3 Dado un punto del test a clasificar, lo añadimos a cada uno de los subconjuntos de entrenamiento separados por clases.
- 4 Recalculamos para cada clase la nueva silueta.
- 5 Calculamos la distancia euclídea entre las siluetas sin el punto de test y las siluetas con el punto del test.
- 6 Asignamos como clase etiquetada aquella cuya silueta haya cambiado menos.

Clasificador topológico

Ejemplo de un conjunto de datos con tres clases diferentes (M_0, M_1, M_2).



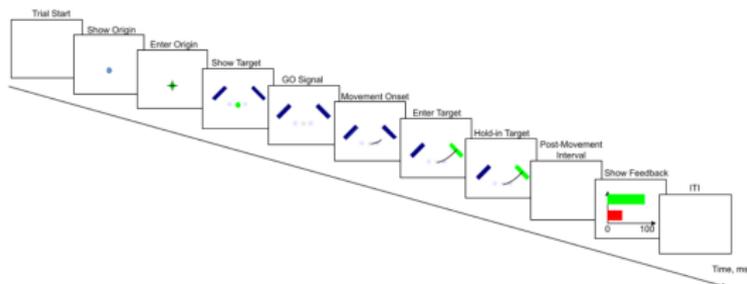
En ambos casos, clasificaríamos con la etiqueta M_0 .

Estructura de la presentación

- 1 Análisis topológico de datos
- 2 Clasificador topológico
- 3 **Detalles del estudio**
- 4 Resultados
- 5 Conclusiones

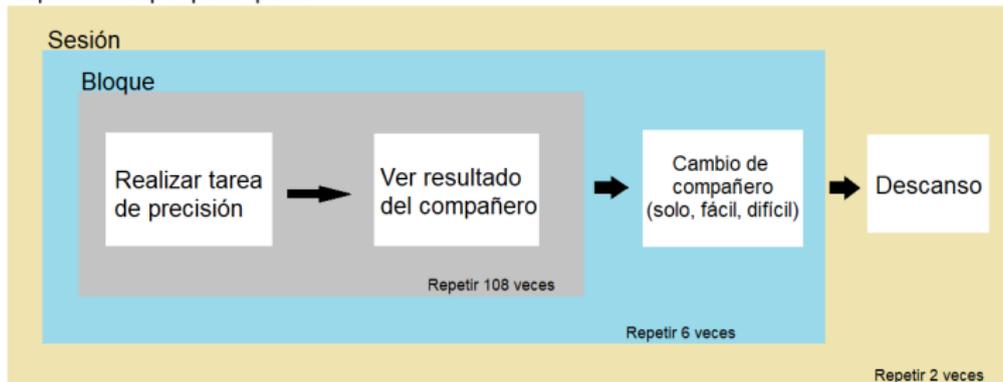
Detalles del experimento

- Se dispone de **64 electrodos** para grabar electro-encefalogramas (EEG).
- Los participantes realizan una tarea de **precisión**.
- Se informa de que la tarea se **realiza con un compañero**, pero **no es competitiva**.
- Al final de cada experimento, se muestra el resultado del participante y el resultado del compañero.
- Se repite el experimento con **tres niveles de presión social**: solo (sin presión), con un compañero de menor habilidad (fácil) y con un compañero de mayor habilidad (difícil).



Detalles del experimento

Experimento por participante



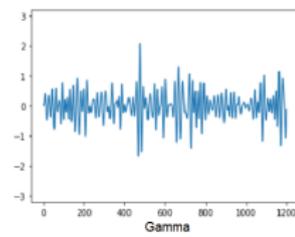
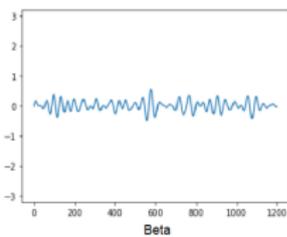
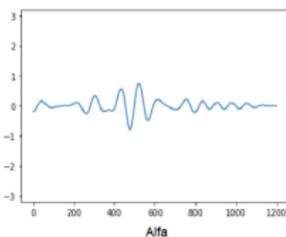
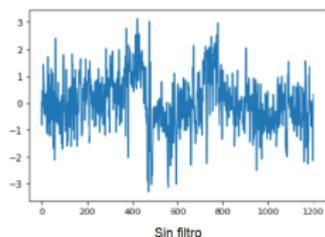
- En resumen: 2 sesiones, cada sesión tiene 6 bloques de experimentos, cada bloque de experimentos consiste en 108 repeticiones de la tarea.
- Para cada participante: 1296 grabaciones de electro-encefalogramas (EEG) de 1200 ms divididas equitativamente para los 3 estados motivacionales diferentes.
- Total: 11 participantes. **Haremos un modelo para cada participante.**

Espacio de fuentes y espacio de electrodos

En este experimento, tratamos los datos desde dos perspectivas diferentes:

- **Espacio de electrodos:** analizamos los EEG obtenidos directamente de los electrodos.
- **Espacio de fuentes:** aplicamos Análisis de Componentes Independientes (ICA) al espacio de electrodos, con el fin de identificar las fuentes reales del cerebro que emiten las señales neuronales.

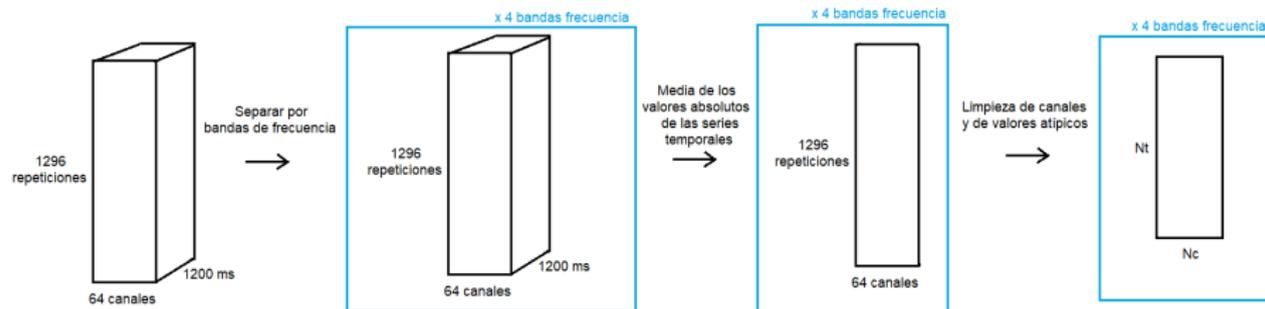
En ambos casos, los EEG serán procesados para estudiarlos según diferentes bandas de frecuencia (sin procesar, alfa, beta y gamma).



Pre-procesamiento

Cada participante representa un conjunto de datos independiente.

- Separamos el EEG en sus diferentes **bandas de frecuencia**.
- Para cada canal, se calcula la **media del valor absoluto** de la serie temporal de su correspondiente EEG.
- **Limpieza** de los canales ruidosos.
- Eliminación de los **valores atípicos**.

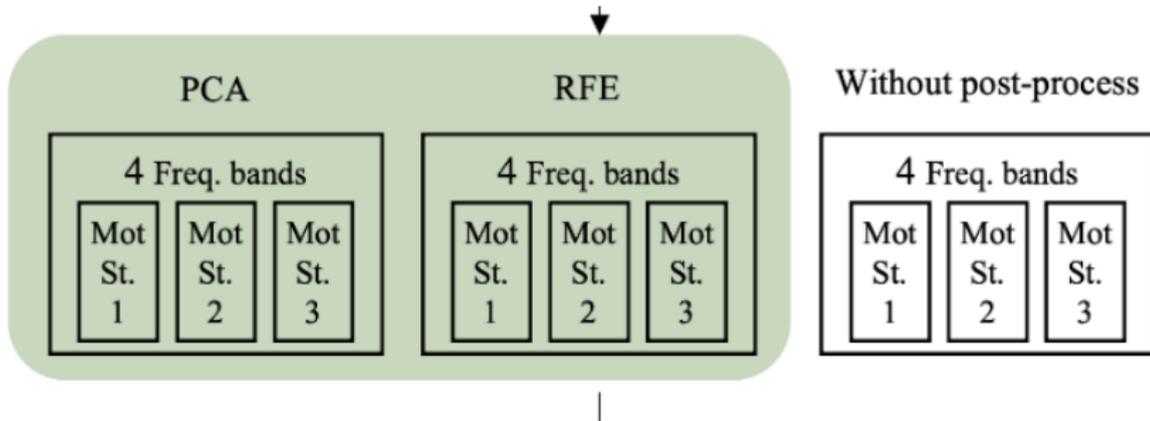


Post-procesamiento

Nos interesa estudiar el **efecto de la dimensión sobre la clasificación topológica**. Para ello llevamos a cabo lo siguiente:

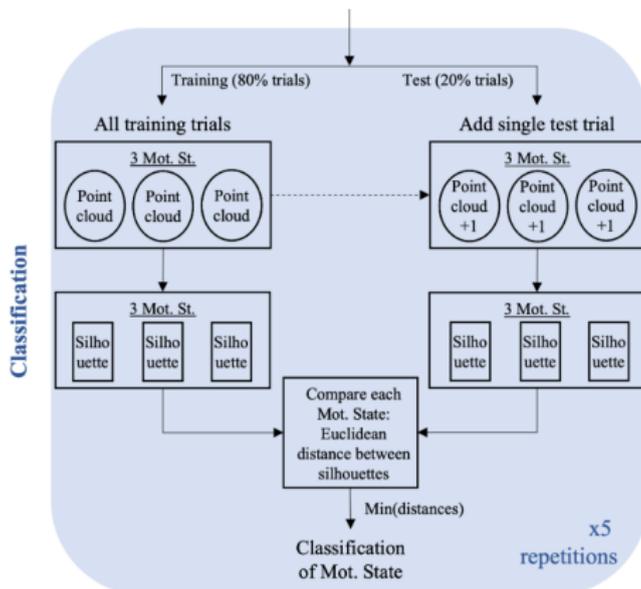
- Aplicar **Análisis de Componentes Principales (PCA)**, calculando de 2 a 10 componentes.
- Usar **Selección Recursiva de Variables (RFE)**, obteniendo de 2 a 10 variables.
- Analizar los datos **sin reducir su dimensión**.

Post-processing



Clasificación

Para cada participante, para cada banda de frecuencia y para cada paso del post-procesamiento, repetimos 5 veces la clasificación y calculamos la precisión.

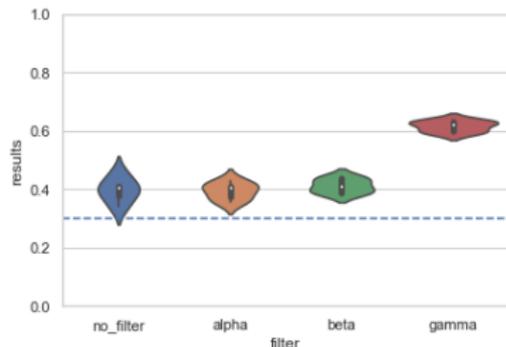
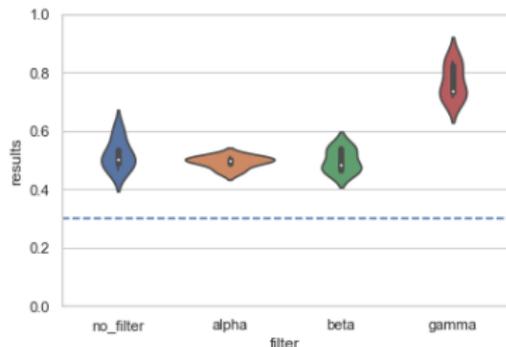
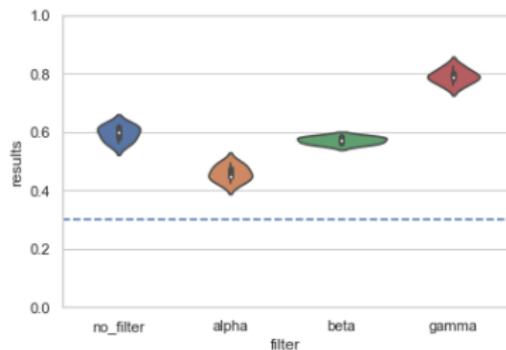
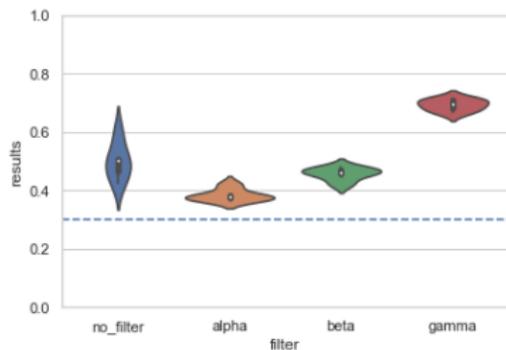


Estructura de la presentación

- 1 Análisis topológico de datos
- 2 Clasificador topológico
- 3 Detalles del estudio
- 4 **Resultados**
- 5 Conclusiones

Resultados base

Precisión de la clasificación de los datos sin reducción de la dimensión para cada banda de frecuencia para los participantes 1, 3, 7 y 11 en el espacio de fuentes.

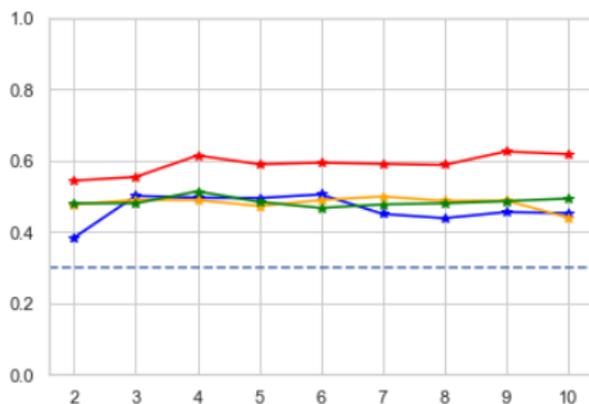
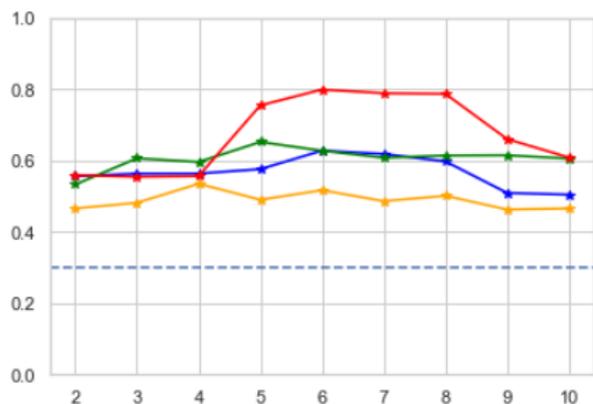


Resultados generales

	no filter			alpha			beta			gamma		
	res	pca	rfe	res	pca	rfe	res	pca	rfe	res	pca	rfe
1	0.50	0.52	0.63	0.38	0.59	0.53	0.46	0.64	0.65	0.69	0.75	0.80
2	0.42	0.45	0.51	0.41	0.52	0.50	0.46	0.56	0.51	0.56	0.59	0.62
3	0.59	0.55	0.70	0.46	0.46	0.45	0.57	0.62	0.58	0.79	0.81	0.74
4	0.44	0.48	0.48	0.41	0.50	0.41	0.48	0.53	0.47	0.50	0.58	0.59
5	0.45	0.58	0.64	0.53	0.66	0.62	0.54	0.56	0.63	0.51	0.61	0.64
6	0.57	0.61	0.70	0.49	0.58	0.57	0.54	0.72	0.63	0.70	0.75	0.72
7	0.52	0.57	0.60	0.50	0.52	0.55	0.50	0.54	0.60	0.77	0.73	0.74
8	0.53	0.55	0.59	0.36	0.53	0.53	0.38	0.67	0.62	0.61	0.73	0.74
9	0.46	0.63	0.72	0.30	0.47	0.46	0.29	0.48	0.51	0.75	0.87	0.81
10	0.34	0.42	0.41	0.35	0.51	0.44	0.30	0.51	0.44	0.42	0.66	0.54
11	0.40	0.41	0.44	0.40	0.43	0.42	0.41	0.49	0.53	0.62	0.70	0.62

Dimensión y RFE

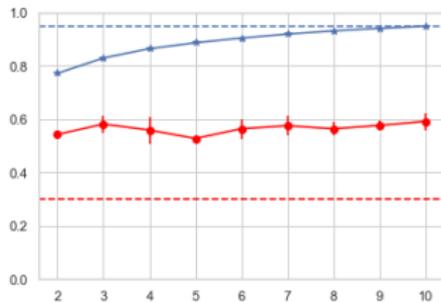
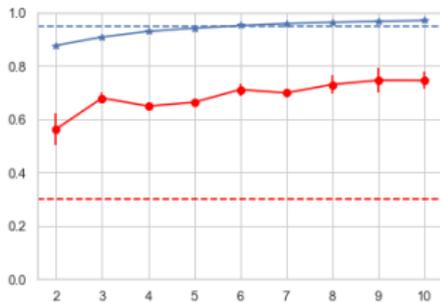
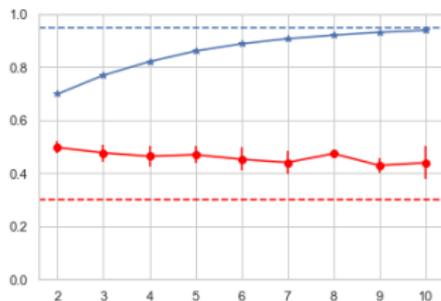
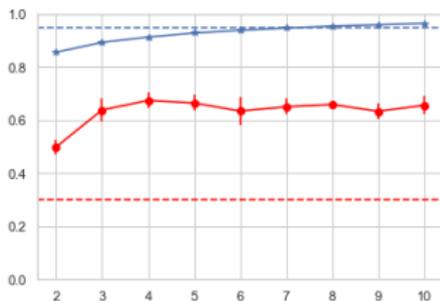
Evolución de la precisión del clasificador topológico según el número de variables seleccionadas por el RFE, participantes 1 y 2, en el espacio de fuentes.



En general, para la banda gamma, 5 variables son suficientes para alcanzar la precisión máxima del clasificador.

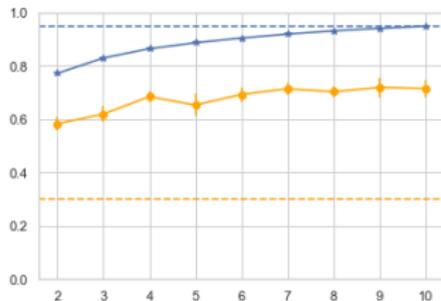
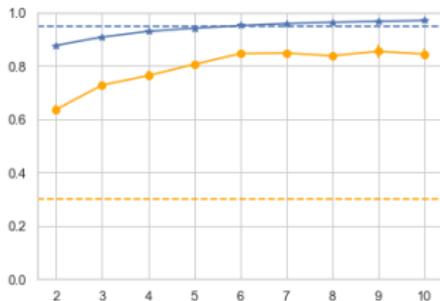
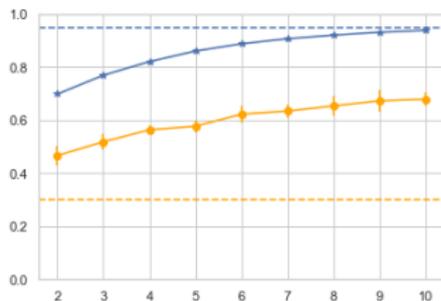
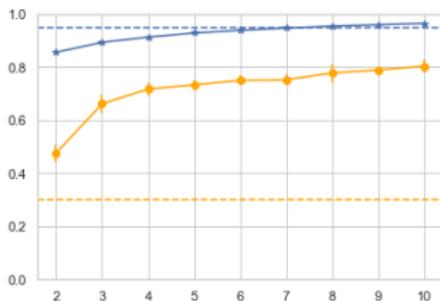
Dimensión y PCA

Evolución de la precisión del clasificador topológico según el número de componentes calculadas en PCA, participantes 1 y 2, en el espacio de electrodos (primera fila) y en el de fuentes (segunda fila).



Comparando con un 1NN

Evolución de la precisión de un clasificador *1-nearest-neighbour* según el número de componentes calculadas en PCA, participantes 1 y 2, en el espacio de electrodos (primera fila) y en el de fuentes (segunda fila).



Estructura de la presentación

- 1 Análisis topológico de datos
- 2 Clasificador topológico
- 3 Detalles del estudio
- 4 Resultados
- 5 Conclusiones

Conclusiones

Por una parte, verificamos que el clasificador topológico es útil para los datos de neurociencia del comportamiento.

Conclusiones

Por una parte, verificamos que el clasificador topológico es útil para los datos de neurociencia del comportamiento.

- El clasificador topológico **obtiene resultados de precisión similares** a otros clasificadores con los datos de este estudio de neurociencia del comportamiento (1-nearest-neighbour).

Conclusiones

Por una parte, verificamos que el clasificador topológico es útil para los datos de neurociencia del comportamiento.

- El clasificador topológico **obtiene resultados de precisión similares** a otros clasificadores con los datos de este estudio de neurociencia del comportamiento (1-nearest-neighbour).
- Usando el clasificador topológico y RFE, **obtenemos conclusiones similares a otros estudios realizados:**
 - 1 Es más fácil clasificar en el espacio de fuentes.
 - 2 La banda de frecuencia gamma es la más útil.
 - 3 Aplicando RFE en el espacio de fuentes vemos que, en general, 5 fuentes son suficientes para la clasificación de estados motivacionales.

Conclusiones

Por otra parte, extraemos conclusiones propias del clasificador topológico que nos dan información extra sobre el problema.

Conclusiones

Por otra parte, extraemos conclusiones propias del clasificador topológico que nos dan información extra sobre el problema.

- La técnicas topológicas que usamos **nos dan una visión complementaria sobre los datos, más allá de su variancia:** el clasificador topológico no está fuertemente relacionado con la variancia explicada.

Conclusiones

Por otra parte, extraemos conclusiones propias del clasificador topológico que nos dan información extra sobre el problema.

- La técnicas topológicas que usamos **nos dan una visión complementaria sobre los datos, más allá de su variancia:** el clasificador topológico no está fuertemente relacionado con la variancia explicada.
- **El espacio de fuentes tiene una forma más complicada:** necesitamos más dimensiones para diferenciar topológicamente los estados motivacionales en el espacio de fuentes que en el de electrodos.

Conclusiones

Por otra parte, extraemos conclusiones propias del clasificador topológico que nos dan información extra sobre el problema.

- Las técnicas topológicas que usamos **nos dan una visión complementaria sobre los datos, más allá de su variancia**: el clasificador topológico no está fuertemente relacionado con la variancia explicada.
- **El espacio de fuentes tiene una forma más complicada**: necesitamos más dimensiones para diferenciar topológicamente los estados motivacionales en el espacio de fuentes que en el de electrodos.
 - ▶ En el espacio de electrodos, el máximo de precisión se obtiene casi siempre en dimensión 4. Podemos decir que **la dimensión de la información obtenida es 4**.

Agradecimientos

Muchas gracias por la atención!



UNIVERSITAT DE
BARCELONA



**UNIVERSIDAD
DE LA RIOJA**

Equipo de trabajo

Carles Casacuberta, Gloria Cecchini, Ignasi Cos, Fritz Pere Nobbe Fisas