Topological analysis in a neuroimaging study

Carles Casacuberta, Ignasi Cos, Aina Ferrà

December 13, 2022

Universitat de Barcelona

Introduction

Reward and motivation are two fundamental drives of human behaviour.

Main objective

Study how the expression of motivation is distributed across the brain.

- Design an experiment focused on identifying the brain network of motivation as modulated by social pressure.
- There is a brain network that conveys social pressure onto a motivational bias, influencing behaviour in the absence of explicit reward.
- **Try to understand the differences between motivational states.**

Introduction

Reward and motivation are two fundamental drives of human behaviour.

Main objective

Study how the expression of motivation is distributed across the brain.

- Design an experiment focused on identifying the brain network of motivation as modulated by social pressure.
- There is a brain network that conveys social pressure onto a motivational bias, influencing behaviour in the absence of explicit reward.
- **Try to understand the differences between motivational states.**

Important

Use **explainable techniques** that help **build up intuition**.

The classification problem

We will deal with a **classification problem**, our goal is to assign labels to our data.

In our experiment

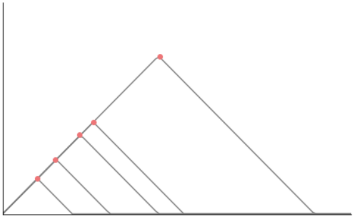
Given some electro-encephalogram data we have to assign to which motivational state it corresponds to.

To test the usefulness of our classifier, we will partition our data in:

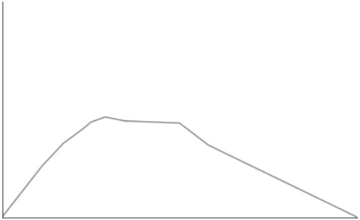
- 80% training, data of which we know the labels.
- 20% testing, data of which we suppose that we do not know the labels and we must assign them.

As an evaluation metric we will compute **accuracy**: percentage of testing data correctly classified.

Topological Data Analysis



Landscape



Silhouette

Tent functions:

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

Landscape functions:

$$\lambda_k(t) = \text{kmax}\{\Lambda_{(b,d)}(t)\}$$

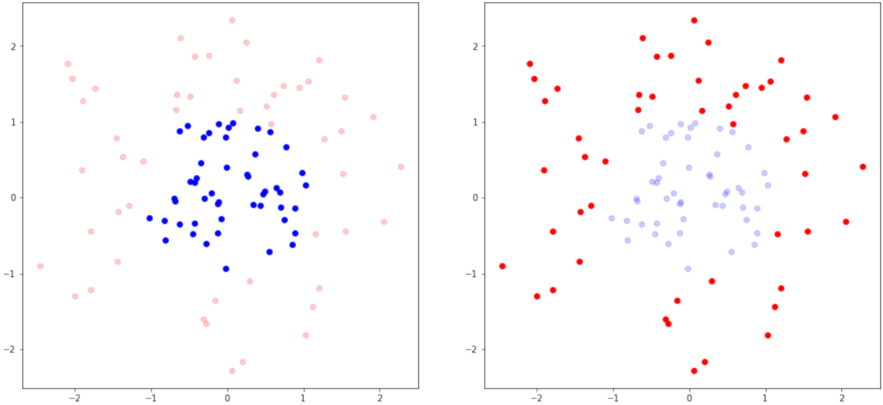
$$\phi_w(t) = \frac{\sum_{i=1}^m w_i \, \Lambda_{(b_i,d_i)}(t)}{\sum_{i=1}^m w_i}$$

Topological classifier

We propose the following algorithm

Persistent based classifier

1. Split the training set into classes according to the given labels.

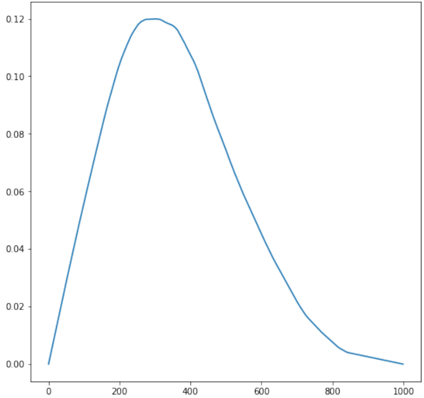


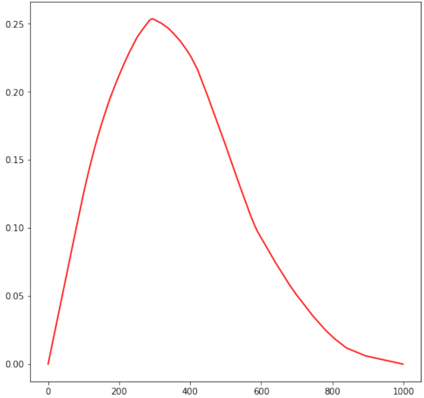
Topological classifier

We propose the following algorithm

Persistent based classifier

2. For each class, compute its silhouette in homological dimension 0.



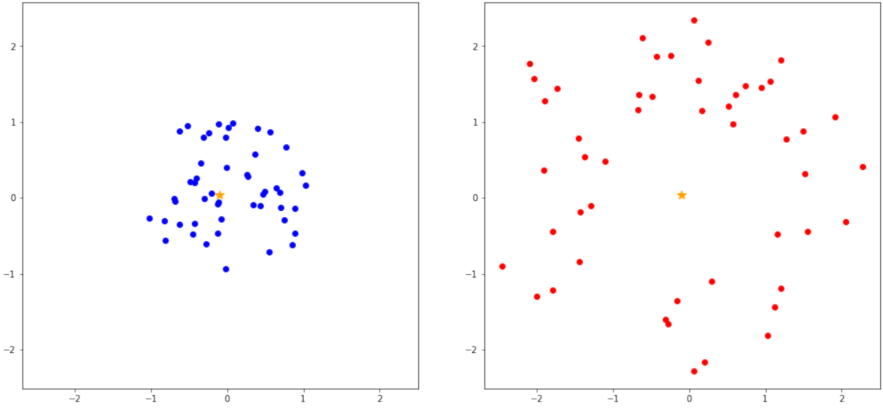


Topological classifier

We propose the following algorithm

Persistent based classifier

3. Given a test input, add it to each of the classes in the training set.

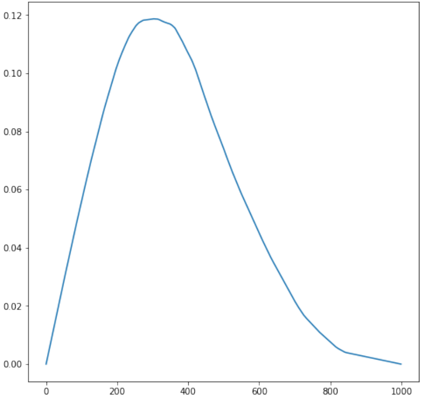


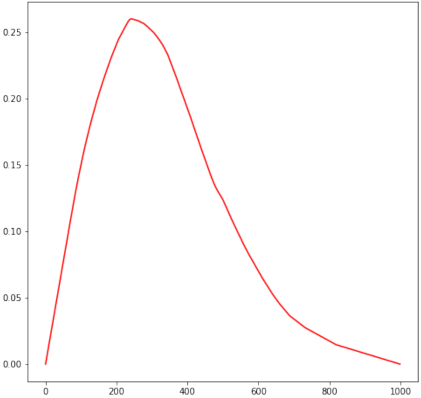
Topological classifier

We propose the following algorithm

Persistent based classifier

4. Recompute the silhouette with the extra point for each class.



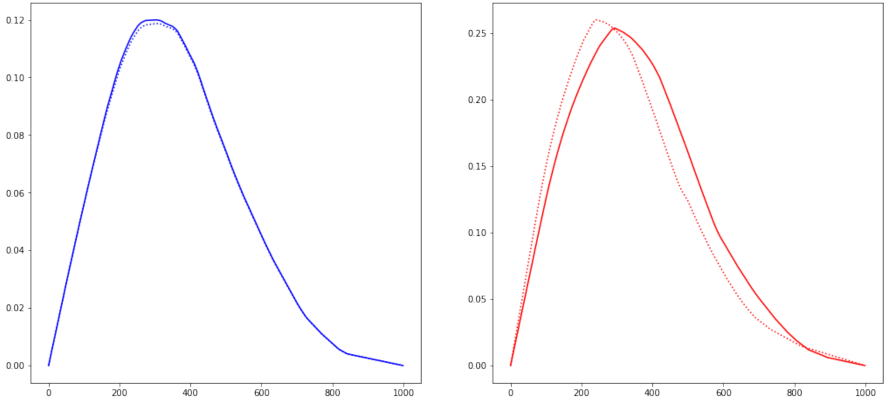


Topological classifier

We propose the following algorithm

Persistent based classifier

5. Calculate the Euclidean distance between the newly obtained silhouettes and the former ones.

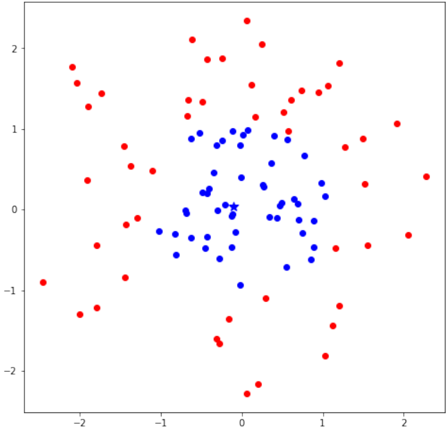


Topological classifier

We propose the following algorithm

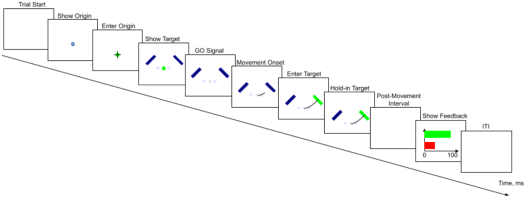
Persistent based classifier

6. Assign as label the class whose silhouette was less perturbed.

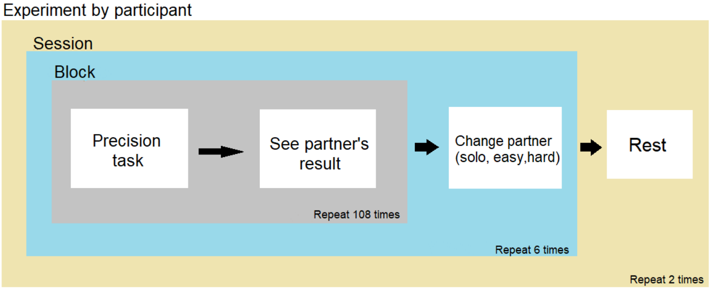


Experiment details

- Use a cap with **64 electrodes** to record electro-encephalograms.
- Participants perform a **precision task**.
- The task is **performed together** with a partner, but it is **not competitive**.
- At the end of each task, the participant's result and their partner's result **are shown together**.
- Repeat the experiment simulating **three motivational states**: solo (**no pressure**), with a less skilled partner (**easy**) and with a more skilled partner (**hard**).



Experiment details

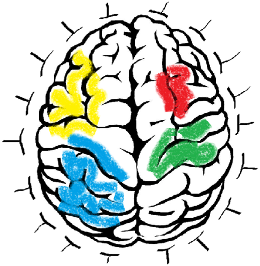


- **In summary**: 2 sessions, 6 blocks of experiments per session, 108 task repetitions per block. A total of 1296 EEG recordings of 1200 ms.
- Data comes from 11 participants. **Each participant has a different model**.
- **Goal:** Given a EEG recording, identify the motivational state.

Space of electrodes and space of sources

Data are looked from two different perspectives.

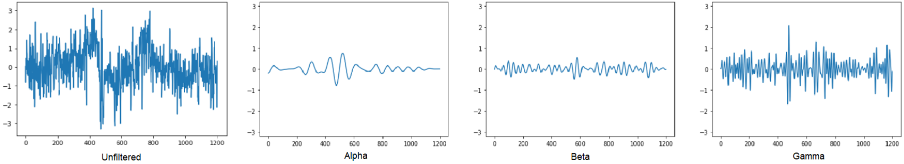
- **Space of electrodes**: EEG data obtained directly from the electrode signal after removing noise and other artefacts.



- **Space of sources**: obtained after applying Independent Component Analysis (ICA) to the space of electrodes with the objective of identifying the real sources of the brain emitting neural signals.

Frequency bands

EEG data is studied by frequency bands: unfiltered (for control purposes), **alpha** (8 - 15 Hz), **beta** (15 - 32 Hz) and **gamma** (32 - 80 Hz), in both the electrode space and the source space.

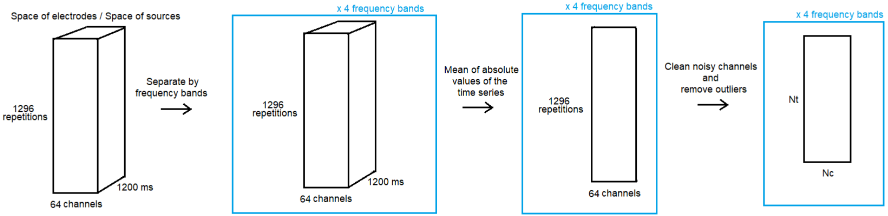


Previous results suggested that **motivation related** modulations would be concentrated on the **higher frequency bands.**

Pre-processing

Data from different participants are not mixed together, so each participant is considered as an independent dataset.

- Separate EEG into the four **frequency bands**.
- For each channel, compute the **mean of the absolute values** of the EEG time series data.
- **Clean** noisy channels.
- **Remove** outliers.

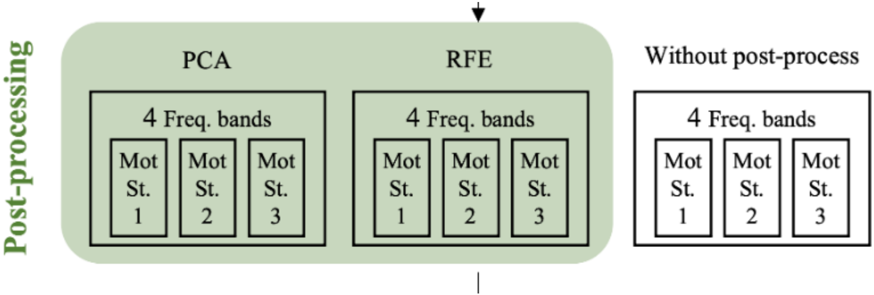


Post-processing

Two different dimensionality reduction methods were applied to the pre-processed dataset, yielding dimensions in the range 2 to 10.

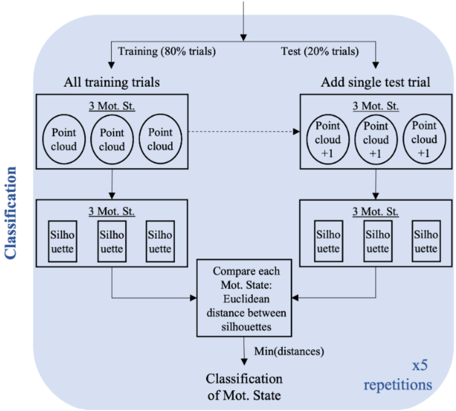
- **Principal Component Analysis**
- **Recursive Feature Elimination**

A version of the data **without dimensionality reduction** was also maintained in order to obtain baseline results.



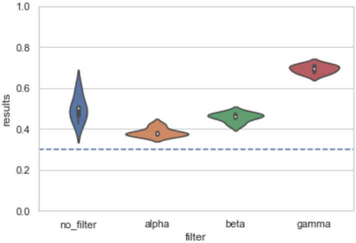
Classification

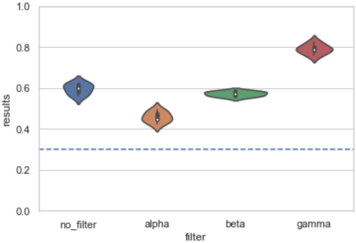
For each participant, frequency band and post-processed dataset, classifications are repeated 5 times and accuracies are averaged over the 5 repetitions.

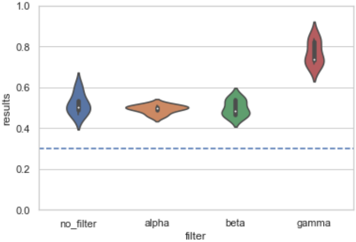


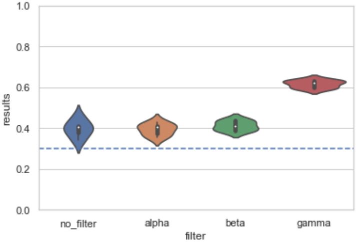
Baseline results

Accuracies of the topological classifier by frequency band using the space of sources and without dimensionality reduction for participants 1, 3, 7 and 11.







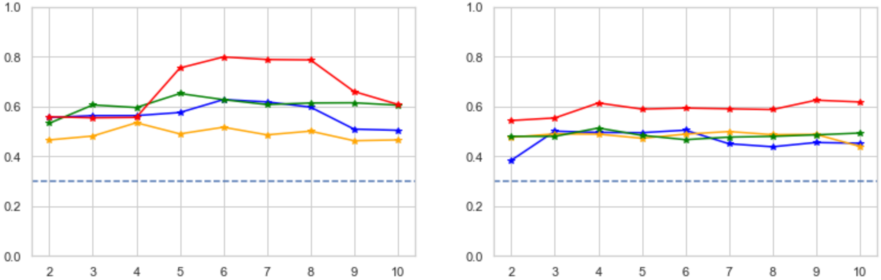


General results

	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

Dimension and RFE

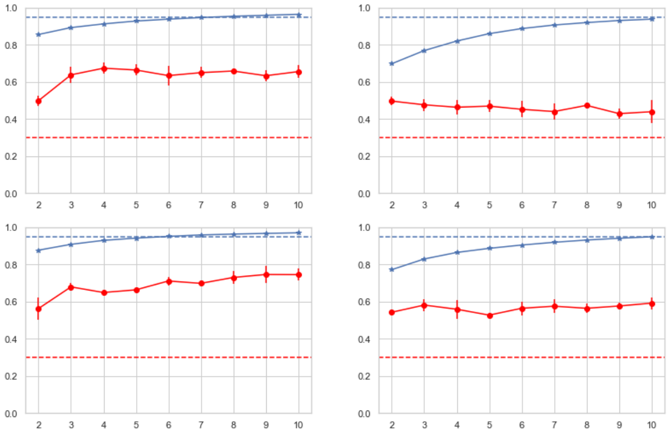
Variation of accuracy of the topological classifier by frequency bands depending on the number of sources selected by recursive feature selection for participants 1 and 2.



In general, for the gamma band, 5 variables are enough to reach a maximum accuracy.

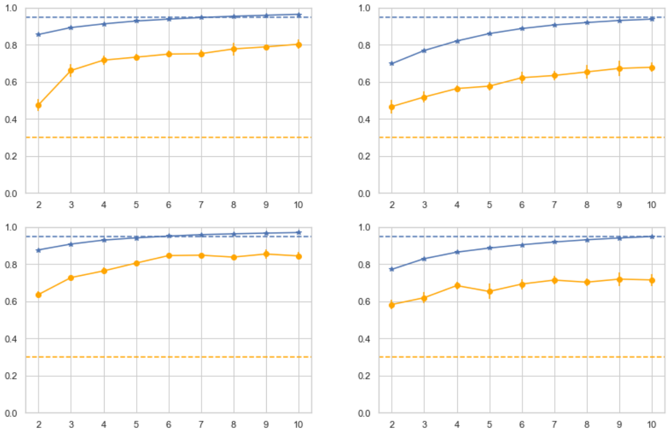
Dimension and PCA

Comparison of variation of accuracy with PCA explained variance as dimension increases for participants 1 and 2 in the electrode space (upper row) and the source space (lower row).



Comparing with a 1NN

Comparison of variation of the accuracy of a 1 nearest-neighbour classifier with PCA explained variance as dimension increases for participants 1 and 2 in the electrode space (upper row) and the source space (lower row).



Conclusions

On the one hand, we find that the topological classifier is useful for behavioural neuroscience studies.

Conclusions

On the one hand, we find that the topological classifier is useful for behavioural neuroscience studies.

- The topological classifier **obtains similar accuracies** as other classifiers with the same data in this behavioural neuroscience study (1-nearest-neighbour).

Conclusions

On the one hand, we find that the topological classifier is useful for behavioural neuroscience studies.

- The topological classifier **obtains similar accuracies** as other classifiers with the same data in this behavioural neuroscience study (1-nearest-neighbour).
- Using the topological classifier together with the RFE technique, we **obtain similar conclusions to ones corresponding to other studies**
  1. The source space contains more information for classification than the electrode space.
  2. The gamma band is the most useful band.
  3. In general, 5 sources are enough to properly classify motivational states.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.
  - The space of sources has a more defined shape due to the ICA algorithm, the maximum accuracy is reached in higher dimensions.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.
    - The space of sources has a more defined shape due to the ICA algorithm, the maximum accuracy is reached in higher dimensions.
- In the case of PCA, **the classifier is more sensitive to the ambient dimension than to the amount of explained variance**.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.
  - The space of sources has a more defined shape due to the ICA algorithm, the maximum accuracy is reached in higher dimensions.
- In the case of PCA, **the classifier is more sensitive to the ambient dimension than to the amount of explained variance**. Explained topology?

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.
  - The space of sources has a more defined shape due to the ICA algorithm, the maximum accuracy is reached in higher dimensions.
- In the case of PCA, **the classifier is more sensitive to the ambient dimension than to the amount of explained variance**. Explained topology?
- The higher sensitive of 1-NN to the explained variance may be due its local nature, as opposed to the global scope of topological summaries.

Conclusions

On the other hand, the topological classifier yields extra information about the problem and opens a door to discuss and investigate new things.

- The accuracy of the topological classifier tends to reach a maximum at low dimensions. In the space of electrodes, the maximum is reached at dimension 4. That suggests that **the latent information contained in our database is close to four-dimensional**.
  - The space of sources has a more defined shape due to the ICA algorithm, the maximum accuracy is reached in higher dimensions.
- In the case of PCA, **the classifier is more sensitive to the ambient dimension than to the amount of explained variance**. Explained topology?
- The higher sensitive of 1-NN to the explained variance may be due its local nature, as opposed to the global scope of topological summaries. Can we improve the topological classifier?