

Computergenerierter Alternativtext:
public 
Files 
private 
Score 

Computergenerierter Alternativtext:
predictions.csv 
predictions.csv 
predictions.csv 
predictions.csv 
0.50007 
0.49735 
0.49735 
0.47149 
0.58085 
0.58402 
0.58402 
0.54857 

Now for this challenge, the dimensionality was manageable, so spatial filtering was not necessary. Still, it is a good thing to reduce dimensionality, so we apply the channel selection procedure to discard irrelevant channels. The number of channels we kept is not very tight (35). it makes everything running more smoothly without having much risk to discard relevant information. EEG has a low spatial resolution, so close channels are highly correlated and often carry the same information. Discarding a few of them doesn't hurt

I think the best book you can find is "Methods of Information Geometry" from S. Amari. The basic idea is that each probability distribution is a point of a manifold with the parameters of the distribution as coordinates. For example a normal distribution of mean u and variance s is a point of coordinate (u,s) in the manifold. The trick is, the Information geometry make the assumption that the natural metric for this manifold is the Fisher information. After a bunch of equation manipulation, it allow you to define a true distance between two distribution, and therefore unlock a lot of issue (for example, the mean of a distribution is not the distribution with the mean parameters :)  )

The data is given to us at 200 samples per second , so each row of the original data corresponds to 5ms. In the  experiment  this data came from an algorithm predicted which letter a person was thinking about, and then showed that letter for 1.3 seconds. We are trying to determine whether the letter shown on the screen is the one the human intended by signals detected with EEG, so I chose data happening when the letter was shown.  260 samples is 5ms\*260 = 1300ms =1.3s.

Looking at couple of the past EEG competitions on Kaggle, it seems most of the effort goes into feature engineering.

For ERP classification, the type and the band of the bandpass filter is not a critical parameters.

An ERP as a typical frequency content that rarely exceed 10 Hz, so the important thing is to cut off the DC value, and remove HF noise. I usually choose to filter between 1 and 20Hz.

As for the type of filter, i suppose IIR are preferred because for an equal order, they are more selective than an equivalent FIR. This could be important for online application (lower order => less lag)

I recommend this paper : *"Interactions Between Pre-Processing and Classification Methods for Event-Related-Potential Classification"*, J. Farquhar · N. J. Hill

The authors studied a lot of different preprocessing methods (including HP cutoff frequency) and their influence on performances.