# fo

# 02460 Advanced Machine Learning LOGBOOK

Rasmus Haarslev, [s152175@student.dtu.dk](mailto:s152175@student.dtu.dk)

Troels Thomsen, [s152165@student.dtu.dk](mailto:s152165@student.dtu.dk)

Marie Le Guilly, [s150497@student.dtu.dk](mailto:s150497@student.dtu.dk)

The main purpose of the logbook is that it serves as a tool for you to organize the project. Further, it serves as a way to collecting information related to the learning objectives:

* Presentation of methods and results at meetings with project supervisor and fellow students
* Plan and carry out the course of the project in collaboration with the project supervisor
* Organize and coordinate the work in the project group

**Overall Project Goals**

**Define own learning objectives for the project**

*The main objective of this project is to clean electroencephalography data (EEG). When recording the activity of a patient’s brain, some high amplitude signal might make the recording noisy: the eye movements, or jaw clenching produce what is referred to as artifacts. Using principal component analysis has been proved successful in removing those artifacts in previous study. Therefore the objective of our project is to reproduce the results of our study.*

*The learning objective are:*

* *Use PCA in a context never encountered before*
* *Get a first experience with brain signals recording*
* *Learn how to choose the parameters of the model*
* *Design experiment to support the parameter choice, and to have performance measures of how well the model is doing.*

**Carry out a well-founded delimitation of the project and formulate specific hypotheses and aims**

*The project should start with simulation. The work has to be carried on a pre-existing dataset for an EEG assumed to cleaned. Then artifacts have to be simulated.*

*If time allows, real time implementation can be done (cleaning while recording from brain), or maybe at least do experiments on data acquired on our own.*

**Project Meetings**

**Week 1: 08.03.2016**

***Questions***

What kind of dataset should we use? Where to find them?

***Reading, who and what***

Website on artifacts removal using ICA

<http://sccn.ucsd.edu/~jung/Site/EEG_artifact_removal.html>

An overview of platforms

Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T.O. and Makeig, S., 2010. MATLAB-based tools for BCI research. In Brain-Computer Interfaces (pp. 241-259). Springer London.

A recent on-line ICA method

Hsu, S.H., Mullen, T., Jung, T.P. and Cauwenberghs, G., 2015. Real-time Adaptive EEG Source Separation using Online Recursive Independent Component Analysis.

***Implementation, who and what***

Tried to install the fieldtrip library which was supposed to interpret eeg data. But finally we found Matlab data, which we transformed into csv files trough a Matlab script we wrote ourselves.

***Results, who and what***

We got csv files of several eeg recording, which is supposed to be clean data. CSV files can be easily used with python.

***Decisions, who and what, what do you don alone, what do you do together***

We decided not to use any of the frameworks we found that day.

**Week 3: 16.03.2016**

***Questions***

What are artifacts? How to simulate them? In which way can PCA helps to remove them?

***Reading, who and what***

Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T.O. and Makeig, S., 2010. *MATLAB-based tools for BCI research.* In Brain-Computer Interfaces (pp. 241-259). Springer London.

***Implementation, who and what***

We made the first version of our code:

* Adding artifacts to an eeg dataset
* Apply PCA to a dataset
* Use PCA results to remove artifacts by projecting onto a subset of components.

***Results, who and what***

We added some artifacts but results were not satisfactory.

***Decisions, who and what, what do you don alone, what do you do together***

We decided to do some work on artifact addition during Easter break.

**Week 4: 31.03.2016**

***Questions***

How can we improve our artifacts? Does using PCA for artifacts removal a good idea?

***Implementation, who and what***

We improved our results by not adding artifacts to all components. We normalized our data prior to PCA and artifacts.

***Results, who and what***

Artifacts seem to be removed. However the price to pay is that cleaned data is sometimes closer to principal components than to original data.

We also installed the libraries that we should use to do live recording with an eeg helmet. But we have not been able to make the emocapper work properly.

***Decisions, who and what, what do you don alone, what do you do together***

We decided to investigate the source of our problems by asking questions during next meeting with our supervisor.

**Week 5: 04.04.2016**

***Questions***

Why are our artifacts not looking like the one in the one in the studies we read?

***Implementation, who and what***

We tried to draw samples from a multi-dimensional Gaussian distribution, with mean and covariance matrix obtain from our dataset. We divided the Gaussian distribution by a factor so that the artifacts obtained by this method would look like spikes.

***Results, who and what***

The shape of our artifacts is still not exactly what we expected, they look too randomly constructed. Moreover we noticed we had some dimensional troubles when we project our data onto the principal components, because we are losing as many dimensions as number of rejected components.

***Decisions, who and what, what do you don alone, what do you do together***

Our priority is now to solve the dimension problem which probably come from a misunderstanding of the PCA projection.

**Week 5: 07.04.2016**

***Questions***

Can we get a correct PCA projection? Is the data we are working on really clean, or are we adding artifacts on already noisy data?

***Implementation, who and what***

We realized we were missing one final step in our PCA procedure: after projecting onto remaining principal components, we add to project our data onto channel space. So we added this last step in our code.

Moreover, we tried to add artifacts onto other dataset we obtained from different studies. We also tried to apply our pca to noisy data obtained from another group who is working also on eeg.

***Results, who and what***

Now that we corrected our PCA procedure, our dimensional troubles are solved. However, even with different datasets, we don’t have the expected results yet: after PCA, the data is flattened and smoothened, but not as much as expected. This could come from two different problems:

* The artifact, which for now are sampled from a Gaussian distribution, still don’t have a satisfying shape.
* Our criterion to reject principal components might not be correct.

***Decisions, who and what, what do you don alone, what do you do together***

We prepared our mid-term presentation, trying to explain well what are problem are, so we could discuss them and maybe get new ideas through discussion.

**Week 6: 14.04.2016**

***Questions***

Does are new solution for artifacts improves our results? How to come with a clever threshold for rejecting principal components?

***Reading, who and what***

Mullen, T.R., et al 2015. *Real-time neuroimaging and cognitive monitoring using wearable dry EEG.* Biomedical Engineering, IEEE Transactions on, 62(11), pp.2553-2567

***Implementation, who and what***

We worked on two problematics:

* Now, our artifacts are added using a sinus function as discussed during the midterm presentation. They look much more like realistic artifacts, with a spike shape.
* We do not always reject components: now, their eigenvalue has to be above some threshold in order for them to be rejected. So we had to find an estimation of the threshold, which we did by looking at the eigenvalues on clean windows of time with no artifacts. With those, we tried two different approaches:
  + Computing the average of those eigenvalues, and deciding to reject components with eigenvalues above 20% more than this average.
  + Looking at the maximum of all eigenvalues for clean windows, and then rejecting components for which the eigenvalue is above this maximum value.

It looks like we have better results with the average value. However we have started to think about more advanced methods such as clustering in order to detect eigenvalues that are not normal.

***Decisions, who and what, what do you don alone, what do you do together***

We know want to try different threshold estimations. To compare them, we will try to measure sensitivity and specificity of the model, by looking at:

* How many times components are rejected in windows on which we did add artifacts
* How many times components were rejected when there should not have been, because we did not have artifacts.

**Week 7: 21.04.2016**

***Questions***

How can we measure how good we are at removing artifacts?

***Implementation, who and what***

We started implementing performance measures. Several approaches were used:

* Specificity and sensitivity are computed by looking at how many times we reject something in windows with artifacts, and how many times we reject when there is actually no artifact.
* Mean squared error between the original dataset and the reconstructed one. We also computed it only on segments with artifacts, and only on segments without artifacts, to see if being good at reconstructed parts affected by artifacts doesn’t impact too much on the part of the signal that were not changed.

Another aspect we’ve been working on is the definition of the threshold. Similarly to the previous week, we had two approaches:

* Using the maximum eigenvalue encountered on the first 10 windows of the dataset
* Using the average of the eigenvalues encountered on the first 10 windows

***Results, who and what***

From the performance measures, it appears that we are probably removing too much after PCA: for many windows with no artifacts, there is still components removed. This is closely related to the threshold, which we could probably improve.

***Decisions, who and what, what do you done alone, what do you do together***

We decided to try new threshold definitions, for example the average of the maximum eigenvalues encountered. Moreover, we need to find a way to adjust our parameters:

* The threshold, obviously
* The window length: we’ve been working with a size of 40, but maybe bigger or smaller windows could improve our results
* The number of window to take into account during the calibration phase that we use to define the threshold. For now we are only working with 10, which is probably too small. A longer calibration should give more accurate results.

**Week 8: 28.04.2016**

***Questions***

Now that we are actually getting some results, and that we have some performance measures, there are deep modifications necessary in our code, in order to make it more coherent with the real experiments our project is aiming at. We need to create a separate process for the calibration phase that would give a threshold that can then be injected in the artifact removal process. Moreover we need to analyze the first performance measures we obtained.

***Implementation, who and what***

The first step has been to reorganize the code in order to be closer to a real life situation simulation. Now, artifacts are added on the dataset before slicing it into windows. This makes more sense because in reality, artifacts might be at the intersection of two windows, and not nicely in the middle of only one.

Moreover, the whole calibration phase, which before was done on the 10 first windows of the dataset prior to artifact removal, has been completed separated into another process. So the calibration program must be run on a “training” dataset, from which a threshold is computed, by applying PCA to the windows of this training test. Then this threshold is injected in the artifact removal process.

We also decided to compare different threshold. The more obvious would be to use the maximum eigenvalue ever encountered in the calibration phase. Anything above it should be considered as abnormal. However, it would be interesting to compare it with other threshold. For example, we could use instead the average of the eigenvalues: this should be way too sensitive, and we risk rejecting too many components, but then we would almost certain to reject the components related to artifacts. Finally, an average of the maximum eigenvalue of each window from the calibration could be a good trade off: it would still be at the upper limit of eigenvalues encountered, but could correct any slightly abnormal value that could be in the calibration phase.

***Results, who and what***

The code is now much clearer, which will make doing experiments much easier. The calibration is also simplified and can be applied really quickly to any dataset.

*Decisions, who and what, what do you don alone, what do you do together*

We want to move faster on the performance measure: mean squared error with cross validation for windows size, sensitivity and specificity to see if artifacts are removed and if artifact free windows are not changed… Moreover we should start on comparing the three thresholds.

**Week 10: 12.05.2016**

***Questions***

What results do we have? What message to we want to deliver with the poster? What amount of information should be displayed on it?

***Implementation, who and what***

As we have now almost all the code we need, and we have a considerable numbers of plots and results, we mainly worked on starting the poster. The main problem was to find a good balance between the amount of text required to explain our work and the limited space available on the poster. Therefore, the plots for the poster have to be carefully selected, and be relevant for the message we want to carry.

***Results, who and what***

A preliminary version of the poster has been done. It contains:

* A short introduction to the problem we are trying to solve
* A mathematical description of the PCA process
* A plot designed like a sort of timeline, to show what we are doing, and the type of result we obtained. It consists in three graphs showing first our original dataset, then a comparison between the original dataset and the same dataset with simulated artifact, and finally a comparison between original dataset and the one reconstructed after PPCA.
* A plot of the results from a cross validation done in order to determine the best window size, for each different threshold we work with. This is useful to explain why we chose this window size for the other plots, and to give a first element of comparison between the different thresholds.
* The sensitivity and specificity measures.
* A conclusion

***Decisions, who and what, what do you don alone, what do you do together***

The design of our poster is not well defined and we should come up with ideas to organize it better during the next meeting. Moreover we need to make sure that we have the right amount of written information.

**Week 11: 16.05.2016**

***Questions***

Is our poster explaining what we’ve done? Should we change thing? Are we sure the results we’re giving are correct?

***Reading, who and what***

*“The Artifact Subspace Reconstruction method”,* Christian A Kothe

***Implementation, who and what***

We checked every result on the poster. We noticed there was a lack of consistency among the results: there were not all based on the same window size nor on the same number and kind of artifacts. We corrected it by rerunning all our experiments and making sure we can give justifications regarding the choice of parameters.

Moreover we considerably changed the design of the poster, so that the structure can help the reader to follow what we are trying to explain: we divided it into several parts that are clearly indicated: purpose, method, results and conclusion. We also agreed on a color code.

Many part of the texts have been removed to obtain a simpler aspect to the poster, assuming we could give additional information while talking during presentation.

Finally we decided who should say what during the presentation.

***Results, who and what***

The general design of the poster is much better. The content has been changed a lot, resulting in less text and more graphics.

***Decisions, who and what, what do you don alone, what do you do together***

Each of us should practice what he has to say for the presentation. Our latest version of the poster has been sent to our supervisor hopping we could get some feedback to improve it if necessary.

**Week 11: 18.05.2016**

***Questions***

How can we improve our poster?

***Implementation, who and what***

Based on the feedback from our supervisor, it appeared that major modifications could still be done in order to improve our poster. The main problem is that there are many thing we didn’t write assuming we could just tell them during presentation. So we had to change our poster in order to write down everything that is necessary for our poster to be “self-contained”.

Moreover we realized we still had some mistakes in our code, resulting in wrong results, which had to be corrected quickly in order to have correct results to show during the presentation. The main problem was our error measures which were biased because we had forgotten to normalize our data.

***Results, who and what***

The poster now contains much more text, and it feels now that someone would really understand what was done by just looking at it and reading the explanations.

Fixing the normalization problem really improved our results, which were then much more coherent.

***Decisions, who and what, what do you don alone, what do you do together***

The only thing left for now is to be ready for the presentation hopping these last modifications of our poster will be enough to comply with what is expected of us.

**Week 11: 22.05.2016**

***Questions***

What should we put in the report?

***Implementation, who and what***

***Results, who and what***

***Decisions, who and what, what do you don alone, what do you do together***

**Supervisor Meetings**

**Week 1: 10.03.2016**

***Presentation of results since last meeting***

We explained the kind of eeg dataset we had found. We also got an eeg recorder, and discussed what kind of programming language we would use: in order to use the eeg helmet, we have to use python with linux.

***Action points for next week***

Try to add artifacts, and to apply a simple PCA, to see if we can manage to remove artifacts.

**Week 2: 17.03.2016**

***Presentation of results since last meeting***

We explained the implementation we had done so far. We asked questions about what artifacts are supposed to look like.

***Action points for next week***

We should try to actually mimic specific artifacts that we can see in the paper he linked us. So one of their artifacts is a sudden dive on the graph, which is the kind of thing we could try to mimic (although what we do right now is not wrong per se). Artifacts occur in time intervals, not constantly.

Removing artifacts: we should just use W instead of Winv because we use PCA and the eigenvectors are orthogonal so it's the same.

We should try plotting the original data (before adding artifacts) and compare it to after we remove artifacts.

**Week 4: 31.03.2016**

***Presentation of results since last meeting***

We explained our methodology for PCA. It appears we were selecting components instead of rejecting them, so we will have to correct this. We should also work on smaller windows of time in order to get more accurate results.

We also solved the small problem we had when we tried to read data from the emocappa: it seems like the helmet and the USB key should be close enough in order for the signal to be detected.

***Action points for next week***

We’re going to apply all the modifications just mentioned in order to see if it improves our results.

**Week 5: 7.04.2016 (Mid-term presentation)**

***Presentation of results since last meeting***

During the presentation, we explained what we had done so far:

* Apply PCA in order to reject components with the highest variance.
* Simulate artifacts, even though their shape is not similar to the one we can see in previous studies.
* Probably because the shape of our artifacts, PCA doesn’t really remove them well.

***Action points for next week***

After discussing our troubles, we found several improvements we could quickly implement for the next week:

* The shape of artifacts should be based on a sinus function (the time being a parameter), with an amplitude function of the channel on which the artifact is injected
* Our criterion used to reject components should be changed: for now, we’ve been working with percentages, rejecting the components above 80% of total variance. However it would be better to work with fixe thresholds, so that when a window doesn’t contain artifacts, nothing would be removed.

**Week 6: 14.04.2016**

***Presentation of results since last meeting***

We explained that our new way of adding artifacts, and using a fixed threshold instead of a percentage considerably improved our results.

***Action points for next week***

We should now come up with performance measures, in order to see how well we actually are at removing artifacts. Moreover, we might have to improve our way of finding the appropriate threshold.

**Week 7: 21.04.2016**

***Presentation of results since last meeting***

We explained all the performance measures we have implemented so far. We also explained that our threshold still needs improvement, and that we are now questioning our choice of parameters.

We also discuss the fact the electrical signal recorded from brains might differ a lot from one person to another, and even from one moment to another. This is why previous to any experiment, a calibration phase will always be necessary to re-compute an appropriate threshold given the conditions.

***Action points for next week***

We have to try to design experiments to adjust our parameters. This could also be a good opportunity to get some visualization of our work, to get some plots that would be the first step of our future poster.

**Week 8: 28.04.2016**

***Presentation of results since last meeting***

We explained all the performance measures we have implemented so far. We also explained that our threshold still needs improvement, and that we are now questioning our choice of parameters.

We also discuss the fact the electrical signal recorded from brains might differ a lot from one person to another, and even from one moment to another. This is why previous to any experiment, a calibration phase will always be necessary to re-compute an appropriate threshold given the conditions.

***Action points for next week***

We have to try to design experiments to adjust our parameters. This could also be a good opportunity to get some visualization of our work, to get some plots that would be the first step of our future poster.