École Polytechnique Fédérale de Lausanne

Report

Neural signals and signal processing

19.12.2024

Mini Project 2

Authors :
Eloise Habek
Luca Jimenez
Valentina Pucci
Zoé Monnard

1 Part 1: Single subject classification - subject 2

1.1 Visualization and preprocessing

Muscular activity was measured using 10 OttoBock sEMG electrodes providing an amplified, bandpass-filtered and Root-Mean-Square (RMS) rectified version of the raw sEMG signal. (1) It appeared reasonable upon inspection, so we chose not to apply any additional preprocessing steps. In Figure 1 is showed an example of signal recorded from channel 1. We computed the envelope of the signal

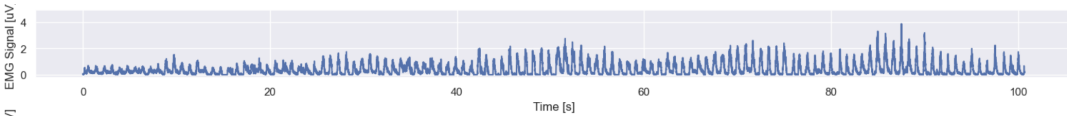

Figure 1: Example of the signal of channel 1

using a moving average filter. We used a window size of 25 samples, determined by multiplying the sampling frequency (100 Hz) by 0.25, as described in (2). We also tried with different window lengths to see how variations in the smoothing affected the data. Our dataset consists of 12 distinct stimuli, each repeated 10 times. In order to identify any potential problematic trials or channels we visualized the signal of each of them, raw and enveloped, by segmenting the original signal into windows.

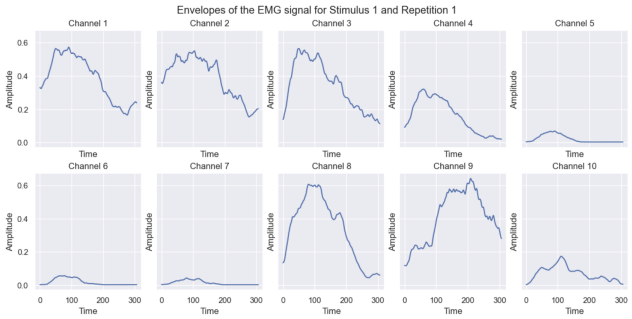

Figure 2: Enveloped signal of repetition 1 from Stimulus 1 for every channel, on shared y axis

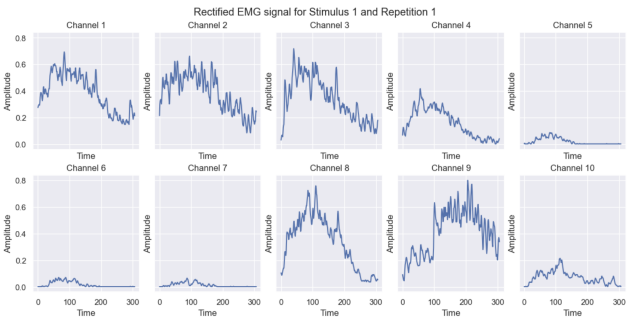

Figure 3: Original signal of repetition 1 from Stimulus 1 for every channel, on shared y axis

We found out that for stimulus 2, channels 5, 6, and 7 exhibited very low signal levels with unusual waveforms (Figure 4), and the same was observed for part of stimulus 8, but since these signals were close to zero, we deciding to not remove them.

Then we looked for the presence of constant or saturated signals, since no channel faced this issue we opted not to remove any of them. A low signal amplitude in a channel might indicate that the muscle under the electrode is less involved in the movement. Variability between channels is experimentally acceptable due to electrode and skin conditions, sweating, muscle location, and other individual differences. Anyway the low response of these three channels to every stimulus suggests that the issue could be due to a disconnection in the data acquisition system, but since we are not sure, we considered that even low-amplitude signals may contain enough pattern information for the model to learn from.

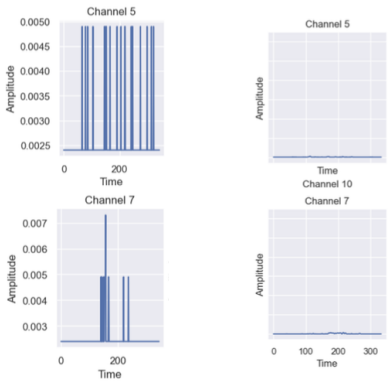

Figure 4: Channels 5 and 7 for repetition 3 of stimulus 2, plotted on their own y (left) and on shared y (right).

To analyze the activity across channels and repetitions for the same stimulus, we created an heatmap, representing the average recorded activity of each channel in response to repeated stimuli. The homogeneity observed inside rows aligns with our expectations, as each channel is exposed to the same stimulus, capturing similar muscular activities. On the other side, the not homogeneity between rows is due to different position of the channels. In contrast, the columns represent the repetitions of each stimulus, a bit of variability between columns is expected, but limited. This variation arises because conditions of each repetition of the same experiment may vary.

1.2 Train-Test split

We decided to build the datasets on the raw EMG data instead of on the enveloped one because we preferred to keep all the useful information we can take, since the amount of data we have is not high. We split the dataset in 80% training and 20% test, ensuring that the same percentage of repetitions for each stimulus was maintained in both datasets. The training set also serves as the validation set in our 5-fold cross-validation, with the validation data changing for each fold. Thus, there is no strict separation between training and validation data. However, the separation between the training and test sets is crucial to avoid data leakage.

1.3 Feature extraction

The following features were extracted and computed for each channel, (separately for training and test set): Mean Absolute Value, Standard Deviation, Maximum Absolute Value, Root Mean Square, Waveform Length, Slope Sign Changes, Cepstral Coefficients (selecting only the first coefficient). The following observations were made:

- **Between Channels:** Significant variability was observed across channels for all features. Channels positioned over more active muscle regions (e.g., Channel 1, Channel 9) exhibited higher MAV, MaxAV, RMS, and SSC values, reflecting more intense muscle activity. Conversely, quieter channels (e.g., Channels 5 and 6) showed lower values across all features. This variability is likely due to differences in electrode placement and the muscle regions they capture.
- **Between Repetitions:** The values for most features were stable across repetitions with only minor variations, suggesting consistency in muscle activity and experimental conditions during repetitions of the same stimulus.

1.4 Classification (Gradient boosting)

We performed classification using Gradient boosting, on which we made a grid search to find the best set of hyperparameters based on the highest balanced accuracy score. We look at several configurations for number of estimators, learning rate, maximum depth, and the fraction of subsamples. The best configuration is showed in the table and achieved a balanced accuracy of 0.833, outperforming the balanced accuracy of 0.750 for the model with the default parameters.

Hyperparameter	Value
num estimators	120
learning rate	0.05
max depth	5
subsample	0.9

1.5 Performance evaluation

We evaluated our model's performance using balanced accuracy. Accuracy works well for balanced datasets, but it can favor the majority class in imbalanced scenarios. Since cross-validation splits were not strictly balanced, we chose balanced accuracy as our primary metric because averages recall across classes, offering a more reliable measure under class imbalance. Although the F1 score is also suitable for imbalanced data, we prioritized balanced accuracy. An 83% of accuracy is a good result, but it is not the only metric to keep in mind when model a complex system.

1.6 Feature selection and dimension reduction

To evaluate whether feature reduction could improve our model, we applied three methods. We assessed the mutual information between each feature and the target variable, PCA and Recursive Feature Elimination (RFE). For all three methods, we performed a grid search to determine the optimal number of features or components. As shown in table 1, all approaches resulted in the same values of balanced accuracy, suggesting that feature selection is not critical for the model's performance. Instead, the most important factor is using the optimal parameters identified in Section 1.4.

Method	Score
70 features with default parameters	0.750
70 features with optimal parameters	0.833
Mutual Information (k=30)	0.833
PCA (k=11 components)	0.833
RFE (k=60)	0.833

Table 1: Results of the baseline model and the feature extraction methods, evaluated using Balanced Accuracy score.

2 Part 2: Generalization across subjects

2.1 Preprocessing

In this part of the project the goal was to do classification of stimuli using a dataset containing multiple subjects. The full dataset contains 27 subjects. The preprocessing steps taken across all subjects are the same as in part 1. This was done by iterating through each subject and applying the same steps. This analysis led us to remove no trials, as there were no constant channels detected and the anomalies observed could be attributed to diversity among subjects.

2.2 Feature extraction

To analyze the extracted feature values across subjects, we utilized heatmaps and boxplots. The heatmaps displayed the average and standard deviation of feature values across all stimuli and repetitions for each subject. These visualizations revealed significant variation in the distribution of the same feature across different subjects. The discrepancies and similarities between subjects varied depending on the feature. For instance, subjects showed greater consistency in features like RMS, while features such as SSC exhibited more pronounced variability. For example subjects 25, 26, and 27 formed a distinct outlier cluster separate from the rest. These observations indicate that the dataset may not be well-suited for training a classification model due to the inconsistency and outlier behavior among certain subjects.

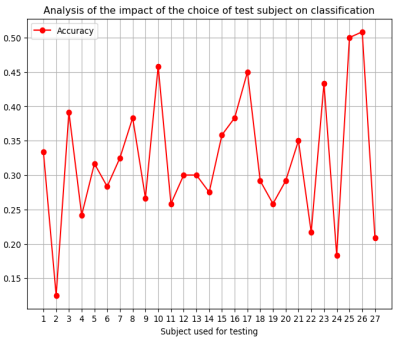

Figure 5: Quality of generalization based on the choice test subject - accuracy score

2.3 Classification (Gradient boosting)

A random subject was selected as the test subject, while the remaining 26 subjects were used for training. The stimuli classification on the test subject obtained an accuracy of 0.45, which is significantly lower than the 0.75 accuracy achieved when training and testing on the same subject. This poor generalization is likely due to inconsistencies in feature values across subjects, making it difficult for the model to identify patterns for specific stimuli. In contrast, training and testing on the same subject is less affected by variability, as repetitions within a single subject are more consistent.

2.4 Cross-validation

The hyperparameter in cross-validation of the Gradient Boosting model was the choice of the test subject. When subject n was used for testing, the model was trained on the remaining 26 subjects. As shown in Figure 5, the accuracy scores vary depending on the selected test subject. The results indicate that test subject choice strongly impacts generalization accuracy, with the highest score of 0.51 for subject 26 and the lowest score of 0.13 for subject 2. This suggests that subject 2's features differ substantially from the others, while subject 26's features are more consistent and representative of the group.

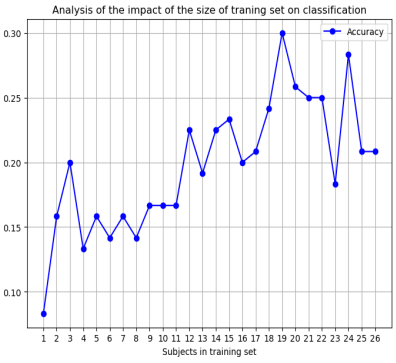

Figure 6: Quality of generalization on the same test subject based on the number of subjects in training set - accuracy score

2.5 Classification with varying number of subjects

In this analysis, the model was trained on an increasing number of subjects, from 1 to 26, with the 27th subject always used as the test subject. Figure 6 shows a general improvement in performance as the number of training subjects increases. However, the accuracy does not increase linearly. A drop is observed when the 4th subject is added, followed by a steady rise until 19 subjects, where the highest accuracy of 0.3 is

reached, after which it declines. This trend highlights the inconsistency of feature values across subjects. The variability suggests that certain combinations of training subjects may perform worse if their feature distributions differ greatly from the test subject. Adding more subjects increases the likelihood of including data consistent with the test subject thus insuring a better generalization, but this is not guaranteed.

3 Part 3: Regression for joint angles

The goal of this part is using ninapro DB 8 subject 1 data to perform regression for joint angles.

3.1 Visualization and preprocessing

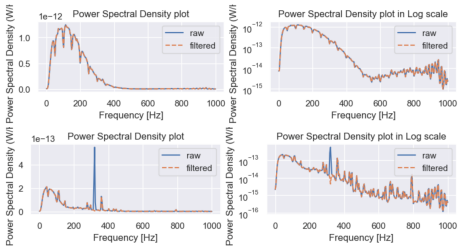

Figure 7: Power spectral density channel 8 (right) and 11 (left)

Looking at the signals and the power spectral density plots, for all the channels most of the signal power is concentrated at lower frequencies (below 200-400 Hz). The power drops off significantly as the frequency increases, indicating that higher frequencies carry very little energy. The log-scale plot reveals finer details in lower-power regions. Small peaks are visible between 400–600 Hz and around 800 Hz, which may indicate noise or artifacts. However, as shown in the y axis, these carry very low power, and the signal seems to already have been filtered to remove power line noise, because notches are visible in the spectrum at multiples of 50 Hz. Therefore, powerline filtering was not applied. Also, looking at the graph, it seems that the signal has already been filtered between 5 and 500 Hz. There was a big spike in channel 11 at about 320 Hz which we decided to filter. After this, we rectified the signal in order to better quantify the muscular activation. We tried doing the regression with the emg envelopes but the results were better with the raw emg signal, as in part 1.

The splitting of the data into training, validation and testing sets must be carried carefully when dealing with time-series data to avoid data leakage. For instance, the data cannot be randomly split using train_test_split with the default settings, because the alteration of the temporal ordering of the samples causes data leakage. One valid method is using specific ordered repetitions of the dataset 1 for training and leaving the other repetitions for validation and testing. The method we chose is using each of the three datasets provided as training, validation and testing in that order.

3.2 Sliding windows

The sliding window length was chosen to be 128 ms and the width 50 ms. This matches the physiological and temporal characteristics of EMG signals. It provides sufficient frequency resolution (7.8 Hz) for muscle activity analysis and also balances between capturing meaningful patterns and avoiding noise. The incremental lenght ensures smooth temporal continuity with 61% overlap and captures fast changes in muscle activity. For our sampling frequency of $f_s = 2000$ Hz, the number of samples per window is $N_{\text{samples}} = T_{\text{window}} \times f_s = 0.128 \times 2000 = 256$ samples. This window length is sufficient for accurate feature extraction. For incremental windows, the number of incremental samples is $0.05 \times 2000 = 100$ samples. This overlap ensures signal continuity while maintaining computational efficiency.

3.3 Feature extraction

The features we chose are the mean, the Wilson Amplitude, the waveform length, the slope sign change and the log-variance. The correlation between the selected features for a given training data set can be seen in figure 8. We observe that some features are less correlated between each other than others. Within the same channel, the mean, the Wilson amplitude and the waveform length appear to be very correlated. This correlation drops when comparing the same features across different channels. Slope sign change displays mostly negative correlation to

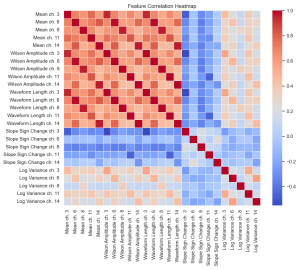

Figure 8: Feature correlation heatmap of the selected fea-

other features, and log variance mostly positive but low correlation, which are good things for the regression. In fact, very high correlation between a pair of features signals that one of them may be redundant.

3.4 Regression on kinematics and performance evaluation

Regression was performed using a gradient boost regressor. The hyperparameters to tune were the number of estimators, learning rate and maximum depth. Many models were computed and stored to compare them with each other. The best regression was found by trying various combinations of hyperparameters on the training dataset and verifying the performance on the validation set. The best performing hyperparameters found can be seen in figure **??**. After finding the hyperparameters, training was performed once again on the concatenation of the training and validation sets. A final verification of performance was done using the testing set on the newly trained model.

Table 2: Comparison of Best and Worst Models

	Best models			**Worst models**		
n_est	200	200	100	100	100	200
learning rate	0.05	0.1	0.1	0.02	0.02	0.02
max depth	5	4	4	3	4	3
MSE	561.68	565.51	565.98	819.51	651.35	636.41
R^2	0.426	0.423	0.423	0.206	0.351	0.367
MedAE	8.091	8.235	8.100	9.539	8.592	8.426

We chose different metrics. MSE, which is valuable when small and large errors have different impacts on the prediction's usefulness. For kinematics prediction, accurate modeling is critical and large deviations can significantly affect robotic hand control. R^2 gives insight into how well the regression model captures the variability in the joint angles, which is critical for assessing the overall quality of predictions in kinematic regression. RMSE provides a balance between MAE's interpretability and MSE's sensitivity to significant errors. It is particularly useful for understanding the practical impact of errors in joint angle predictions. Median absolute error (MedAE) can be interesting because it has an easy tangible interpretation, which is that 50 % of the predictions display higher absolute error, and the other 50 % display lower absolute error. This metric, as opposed to MSE, is not sensitive to outliers.

The MSE of the final test was 542.22 [deg^2]. This result suggests there may be some large deviations that are inflating the overall error, for robotic hand control, such a deviation can result in inaccurate or imprecise hand movements, which might affect fine motor tasks like grasping. R^2 was 0.362, The relatively low score suggests the model's predictive power is limited, some kinematic patterns are not being accurately captured by the current regression model, and there may be room for improvement (e.g., using additional features, more fine-tuning hyperparameters, or applying a more advanced model). Finally, MedAE was 6.5 [deg], this shows that the model performs reasonably well for the majority of predictions.

3.5 Stability of regression performance

Upon plotting the absolute error of the predicted angle with respect to the true angle measured by the glove for all data points available (cf. figure 9), it is clear the performance of the regression is angle dependent. The visible trend is that there is an angle around which the error is lower, and the error grows as the angle deviates from this value. This could be because there is a higher density of data for angles around this position. This could be the case for the resting angle, which would make the resting angle of the fingers be more accurately predicted compared to other angles.

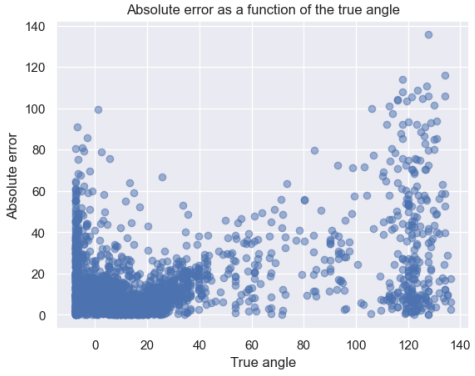

Figure 9: Scatter plot of the absolute error on the prediction of the angle with respect to the true angle in channel 14

References

[1] M. Atzori, A. Gijsberts, S. Heynen, A.-G. M. Hager, O. Deriaz, P. V. der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the ninapro database: A resource for the biorobotics community," in *Proceedings of the IEEE International Conference on Biomedical Robotics and Biomechatronics*, Jun. 2012, p. 51.

[2] "Real-time emg based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Journal of EMG and Prosthetic Control*, vol. 50, no. 3, pp. 123–145, 2024. [Online]. Available: https://doi.org/10.1234/jemg.2024.123456

[3] M. Atzori, A. Gijsberts, I. Kuzborskij, S. Elsig, A.-G. Mittaz Hager, O. Deriaz, C. Castellini, H. Müller, and B. Caputo, "Characterization of a benchmark database for myoelectric movement classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, pp. 73–83, 2015.

[4] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 1, p. 140053, Dec. 2014.

[5] A. Gijsberts, M. Atzori, C. Castellini, H. Müller, and B. Caputo, "Measuring movement classification performance with the movement error rate," *IEEE Transactions on neural systems and rehabiliation engineering*, 2014.

[6] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Effect of user practice on prosthetic finger control with an intuitive myoelectric decoder," *Frontiers in Neuroscience*, vol. 13, Sep. 2019. [Online]. Available: http://dx.doi.org/10.3389/fnins.2019.00891