Rejection-based EEG Signal Classification

Artyom Matveev[1]

matveev.as@phystech.edu

[1]Moscow Institute of Physics and Technology

ASCOMP 2024

Problem Definition and Contribution

Goal: Find an optimal sample such that it includes a minimum amount of information needed to solve an EEG signal classification problem robustly.

Key Contributions: A new approach to the classification of EEG signals.

- First, we reconstruct the probability density function of each class, taking the Riemannian Gaussian distribution of data into account [1].
- Second, we define a specific confidence interval for each class so that we can use it in our rejection strategy.
- Third, we solve the classification problem by evaluating the statistical significance of data concerning the classes' distributions.

Formulation

Assumption: There is an optimal sample size that is enough to make a robust decision during the classification of signals.

1. Let $\mathfrak{D} = \{(\mathbf{X}_i, y_i)_{i=1}^{L}\}$ be the given dataset, where a segment of EEG signals
$$\mathbf{X}_i = [\chi_1, \dots, \chi_j, \dots, \chi_{n_C}]^{\intercal}, \mathbf{X}_i \in \mathbb{R}^{n_C \times n_T} \text{ and } j \in \mathcal{J} = \{1, \dots, n_C\}.$$
2. Let $\mathbf{P}_i \in \mathbb{S}_{++}$ be a Symmetric Positive Definite (SPD) matrix, where
$$\mathbf{P}_i = \frac{1}{n_T - 1}\mathbf{X}_i\mathbf{X}_i^{\intercal}$$
3. Let $y_i \in \mathbb{Y} = \{1, \dots, K\}$ be a class label.
4. Let $p(y|\mathbf{P}, \mathbf{w})$ be a parametric family, where $\mathbf{w} \in \mathbb{R}^n$.
5. The likelihood function is then as follows:
$$L(\mathfrak{D}, \mathbf{w}) = \prod_{i=1}^{m} p(y_i|\mathbf{P}_i, \mathbf{w}), \quad l(\mathfrak{D}, \mathbf{w}) = \sum_{i=1}^{m} \log p(y_i|\mathbf{P}_i, \mathbf{w})$$

Find:

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w} \in \mathbb{R}^n} L(\mathfrak{D}, \mathbf{w})$$

Criterion:

ROC AUC

Method

Sufficient Sample Size Estimation. Bootstrap:

Given some sample size m calculate the quantile confident intervals

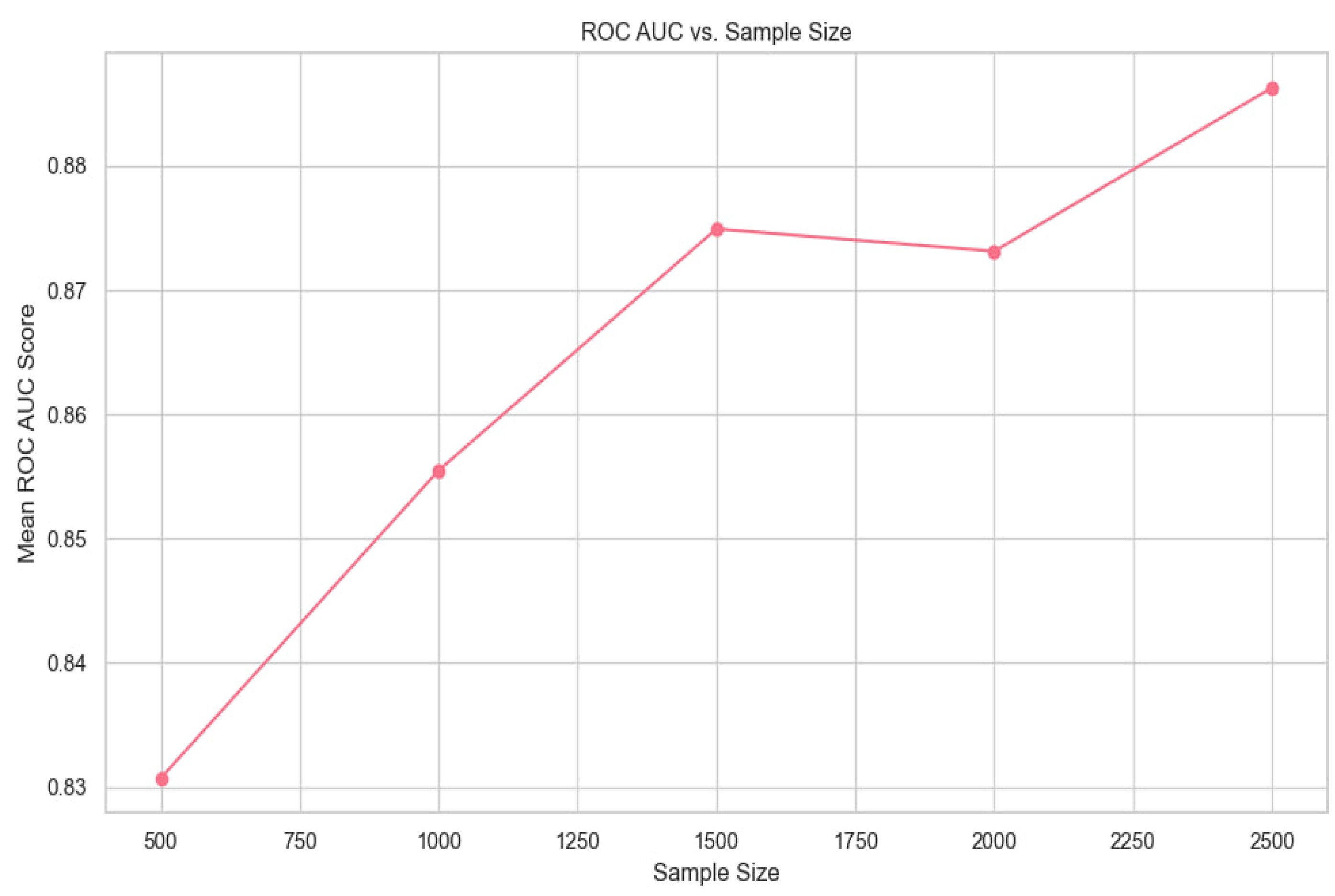
$$(a_1^m, b_1^m), (a_2^m, b_2^m), \dots, (a_n^m, b_n^m)$$

with the significance level of α using bootstrap for every parameter of the model.

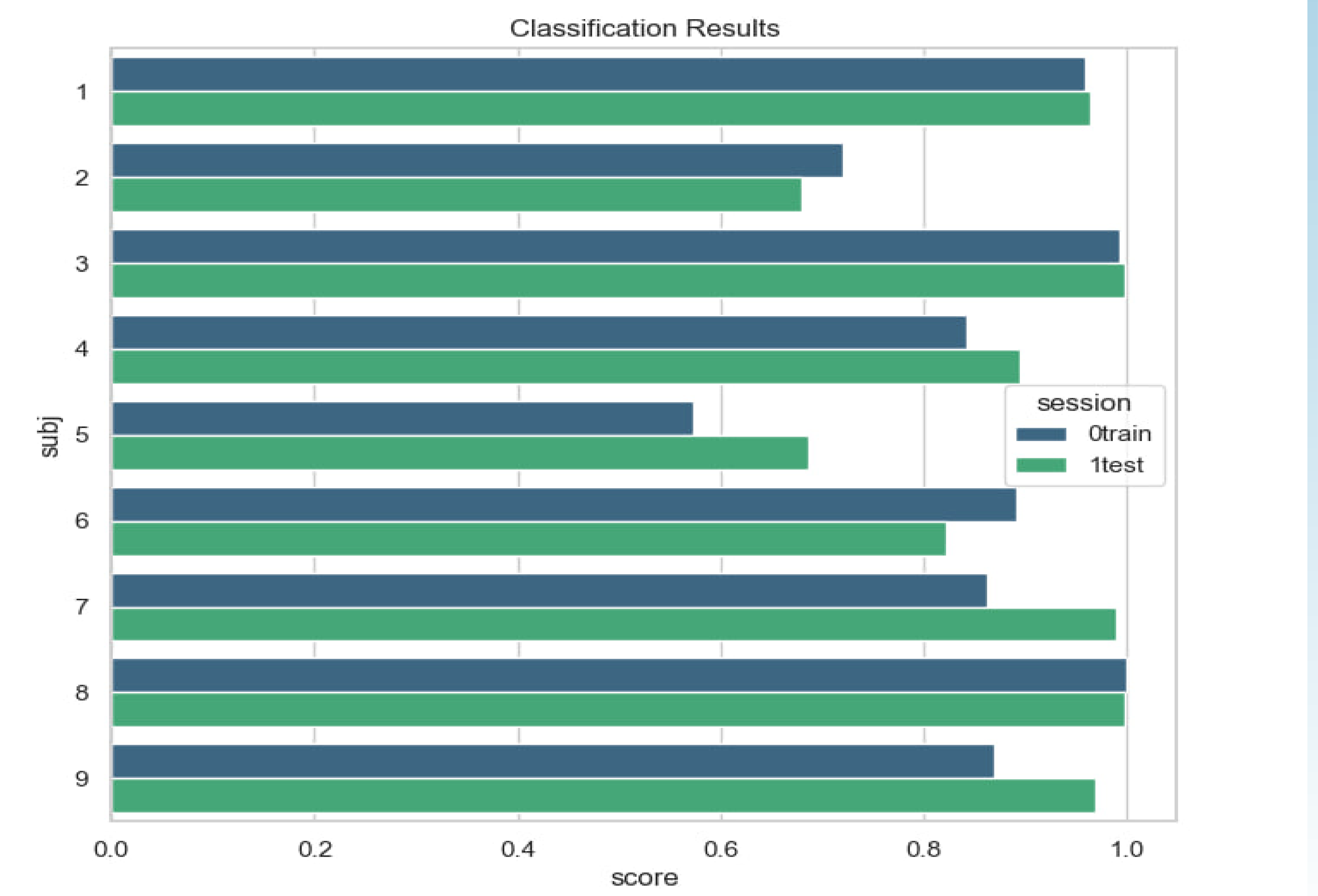
Sufficient sample size $m^*: \forall m \geq m^* \max_i (b_i^m - a_i^m) < l$, where (a_i^m, b_i^m) is a quantile bootstrap confident interval calculated on the i-th bootstrap subset of the m size [2].

Experiments & Results

Quantitative Results. ROC AUC vs Sample Size:



Quantitative Results. ROC AUC Score Given 2500 Examples in the Sample:



References

[1] S. Said, S. Heuveline, and C. Mostajeran, "Riemannian statistics meets random matrix theory: Toward learning from high-dimensional covariance matrices," *IEEE Transactions on Information Theory*, vol. 69(1), 2023.

[2] A. Grabovoy, T. Gadaev, A. Motrenko, and V. Strijov, "Numerical methods of sufficient sample size estimation for generalised linear models," *Lobachevskii Journal of Mathematics*, vol. 43, 2022.