Weakly Supervised Bayesian Estimation of Seizure Likelihood

Noam Siegel

Supervised By
Prof. Oren Shriki & Dr. David Tolpin

Ben-Gurion University of the Negev

January 11, 2023

Outline

Background

Figure 1: 18th century iconograph of an epilepsy seizure. Etched by J. Duplessi-Bertaux. Credit: Wellcome Library, London. Licensed under the Creative Commons Attribution 4.0 International license.

Background

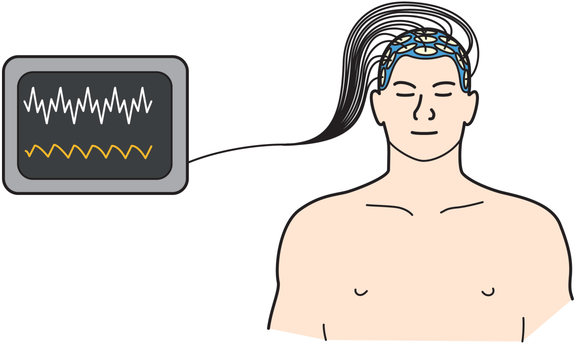
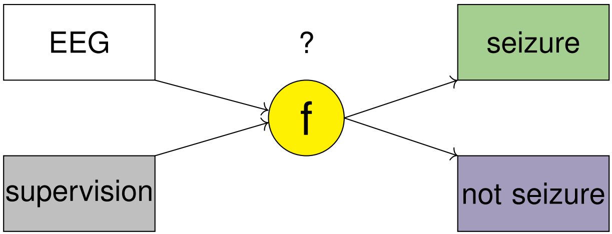
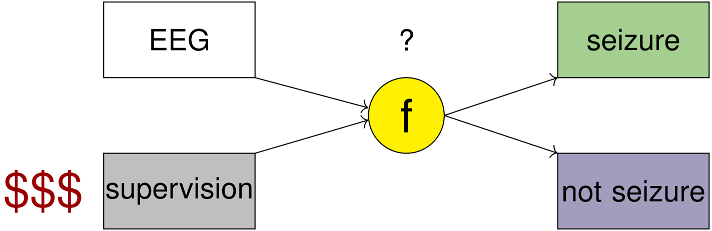


Figure 1: Electroencephalography (EEG) art.
From: The Clear Communication People. Licensed under: CC BY-NC-ND 2.0.

Background & Motivation

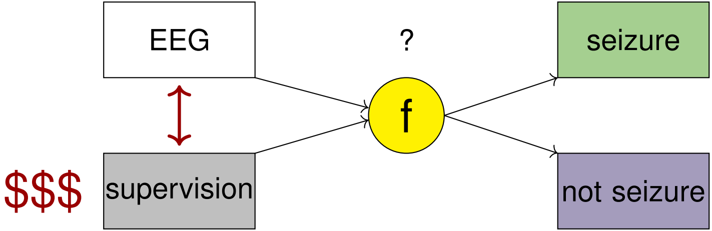


Background & Motivation



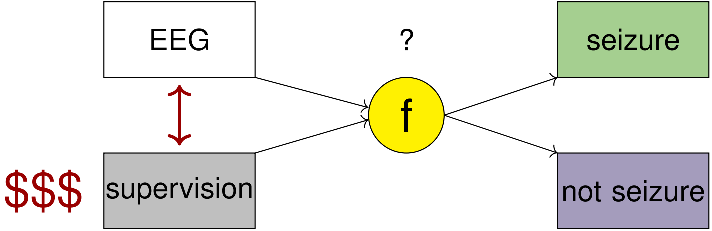
1. Expensive to produce.

Background & Motivation



1. Expensive to produce.
2. Variables are tightly coupled.

Background & Motivation



1. Expensive to produce.
2. Variables are tightly coupled.

Related works

1. Gardner et al. [2006] - unsupervised seizure detection.
2. Karoly et al. [2017] - circadian rhythms as priors.
3. Yang et al. [2022] - weak self-supervised learning.

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.
2. Demonstrating the approach on real-world data, achieving 0.88 AU-ROC with zero labels on a seizure detection task.

Main contributions

Contributions

1. Proposing a Bayesian approach to estimating seizure likelihood using EEG.
2. Demonstrating the approach on real-world data, achieving 0.88 AU-ROC with zero labels on a seizure detection task.
3. A weakly supervised version which is biased towards circadian rhythms is shown to improve detection in canines.

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

Solution

$$\underbrace{\mathbb{P}(S_t \mid E_t)}_{\text{probability of seizure given EEG}} = \frac{\mathbb{P}(S_t)\mathbb{P}(E_t \mid S_t)}{\mathbb{P}(E_t)} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Normalizing factor}}$$

Research problem

Part one

How can we formulate the task of seizure detection as a Bayesian inference problem?

Solution

$$\underbrace{\mathbb{P}(S_t \mid E_t)}_{\text{probability of seizure given EEG}} = \frac{\mathbb{P}(S_t)\mathbb{P}(E_t \mid S_t)}{\mathbb{P}(E_t)} = \frac{\text{Prior} \cdot \text{Likelihood}}{\text{Normalizing factor}}$$

Part two

How can we compute these values?

Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

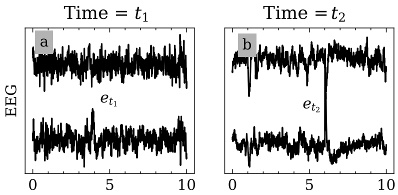


Figure 2: Samples from $E(t_1)$ and $E(t_2)$.

Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

and:

2. $s_t \sim S(t) \in \Omega_S = \{0, 1\}$

a random Seizure variable.

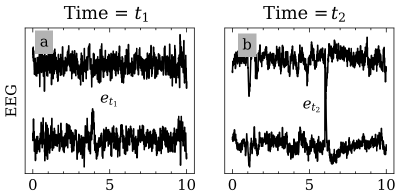


Figure 2: Samples from $E(t_1)$ and $E(t_2)$.

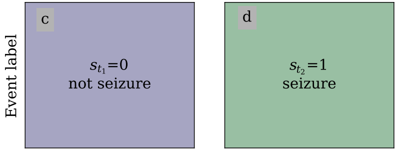


Figure 3: Samples from $S(t_1)$ and $S(t_2)$

Highest Density Regions

Adapted from Hyndman [1996]:

Definition (α-Highest Density Region - α-HDR)

Let $p(e)$ be the density function of a random variable E. For a fixed $0 < \alpha < 1$, and a sample space Ω_E:

$$L_\alpha\text{-HDR} := \{e : p(e) \geq L_\alpha\} \subseteq \Omega_E$$

where

$$L_\alpha \equiv \sup_{\mathbb{R}}\{L : \mathbb{P}(e \in L_\alpha\text{-HDR}) \geq 1 - \alpha\}$$

Highest Density Regions

Adapted from Hyndman [1996]:

Definition (α-Highest Density Region - α-HDR)

Let $p(e)$ be the density function of a random variable E. For a fixed $0 < \alpha < 1$, and a sample space Ω_E:

$$L_\alpha\text{-HDR} := \{e : p(e) \geq L_\alpha\} \subseteq \Omega_E$$

where

$$L_\alpha \equiv \sup_{\mathbb{R}}\{L : \mathbb{P}(e \in L_\alpha\text{-HDR})) \geq 1 - \alpha\}$$

Corollary

$$Vol(L_\alpha\text{-}HDR) \equiv \inf_{R \subseteq \Omega_E} \{Vol(R) : \mathbb{P}(e \in R) \geq \alpha\}$$

Theoretical Basis

Definition (Normal model - $\not{S}$)

For a fixed $0 < \alpha < 1$, a model for normal EEG in Ω_E is:

$$\not{S} := L_\alpha\text{-HDR} \subseteq \Omega_E$$

Theoretical Basis

Definition (Normal model - $\not{S}$)

For a fixed $0 < \alpha < 1$, a model for normal EEG in Ω_E is:

$$\not{S} := L_\alpha\text{-HDR} \subseteq \Omega_E$$

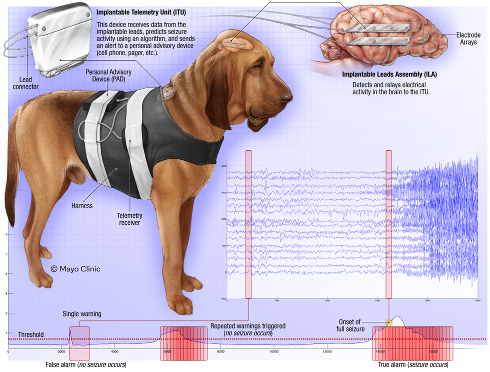
Research problem

$$\mathbb{P}(S_t \mid E_t) = 1 - \mathbb{P}(\not{S}_t \mid E_t) = 1 - \frac{\mathbb{P}(E_t \mid \not{S}_t)\mathbb{P}(\not{S}_t)}{\mathbb{P}(E_t)}$$

1. $\mathbb{P}(e_t)$ is the likelihood of observing e_t in general.
2. $\mathbb{P}(e_t \mid \not{S}_t)$ is the likelihood of observing $e_t \in \not{S}_t$.
3. $1 - \mathbb{P}(\not{S}_t)$ is the prior belief of a seizure at time t.

How do we calculate these values?

Data



Canine Epilepsy Dataset

Figure 4: From Coles et al. [2013]. With permission of Mayo Foundation for Medical Education and Research, all rights reserved.

Longitudinal data is scarce.

Invasive systems offer high SNR.

4 canines,
median 390 days iEEG,
made publicly available by NINDS.

We use 1 canine with 475 days iEEG.

Evaluation plan - non-informative prior

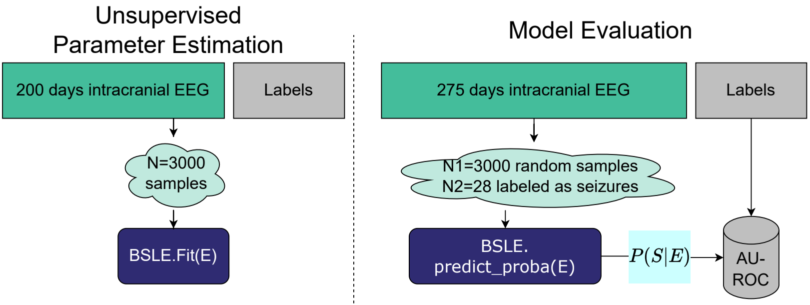


Figure 5: Simulating a Prospective Study - Unsupervised Model

Evaluation plan - informative prior

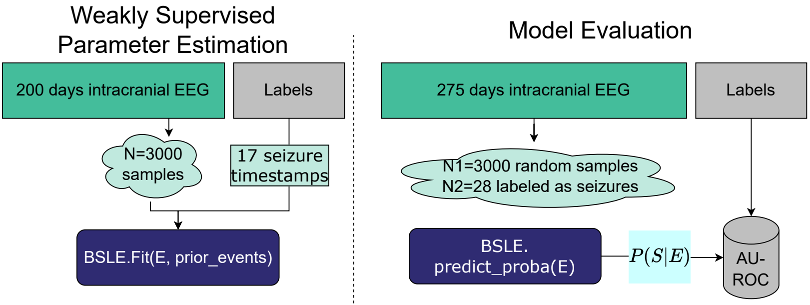


Figure 6: Simulating a Prospective Study - Weakly Supervised Model

Methods

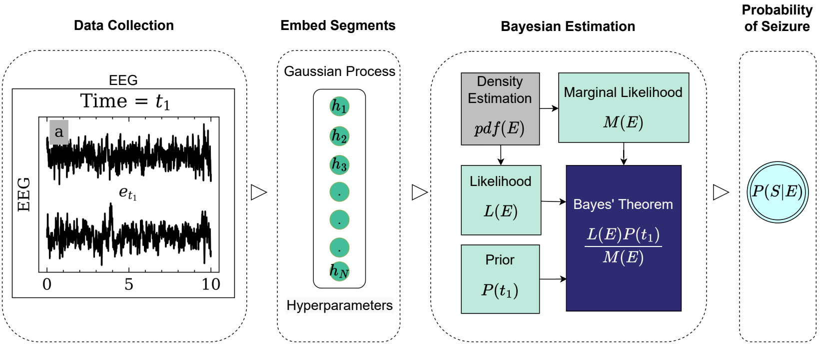


Figure 7: Bayesian Seizure Likelihood Estimation (BSLE) pipeline

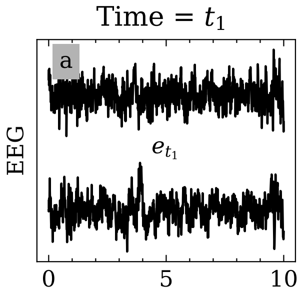
Embedding an EEG segment, by

maximizing the likelihood of:

$$e(t_1) \sim \mathcal{GP}(m(t), k(t, t'))$$

$$m(t) := \mathbb{E}[e(t)]$$

$$k(t, t') := \mathbb{E}[(e(t) - m(t))(e(t') - m(t'))]$$



subject to:

- $e(t_1)$ is a given z-scored EEG segment.
- $m \equiv 0$.
- $k \equiv K_{Matern}(\frac{3}{2})$.

Computed embeddings

$$\underset{d=8000}{e_{t_1}} \xrightarrow[V.I.]{} \underset{d=8}{h_{t_1}}$$

For each sample, estimate:

- signal covar. K_{data}
- task covar. K_{tasks}
- obs. noise $\vec{\epsilon}$

parameters which maximize:

$$p(f(e_{t_1}) \mid h_{t_1})$$

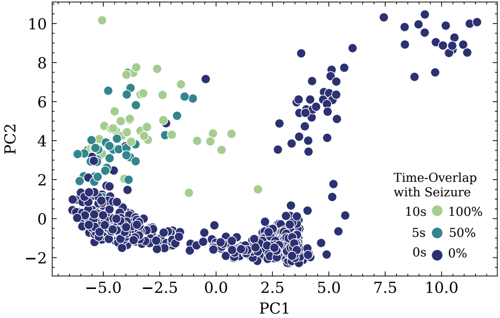
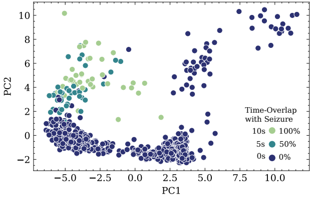


Figure 8: EEG embeddings. Eight dimensional embeddings preserve class separability.

Calibrating Internal Model Parameters



Embeddings

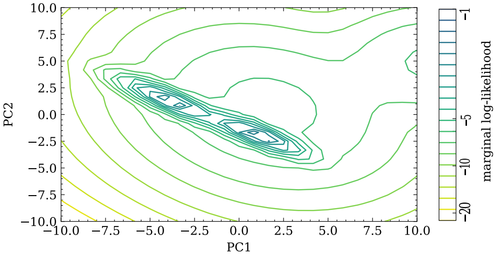


Figure 9: Estimated $\hat{p}(h)$ using Gaussian mixture model

Model of Normal EEG

Definition (Normal model - $\not{S}$)

For a fixed $0 < \alpha < 1$, a model for normal EEG in Ω_E is:

$$\not{S} := L_\alpha\text{-HDR} \subseteq \Omega_E$$

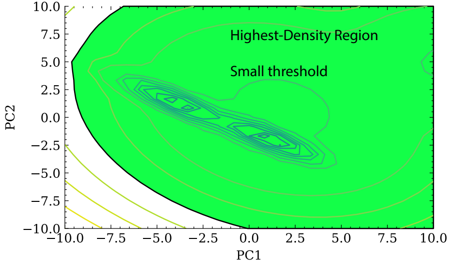


Figure 10: Illustration of $\not{S}$ with small α.

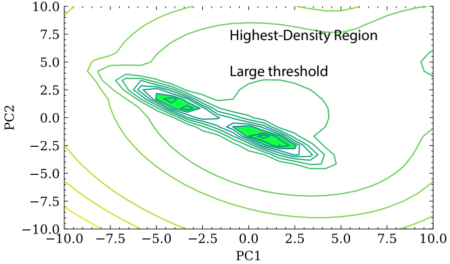
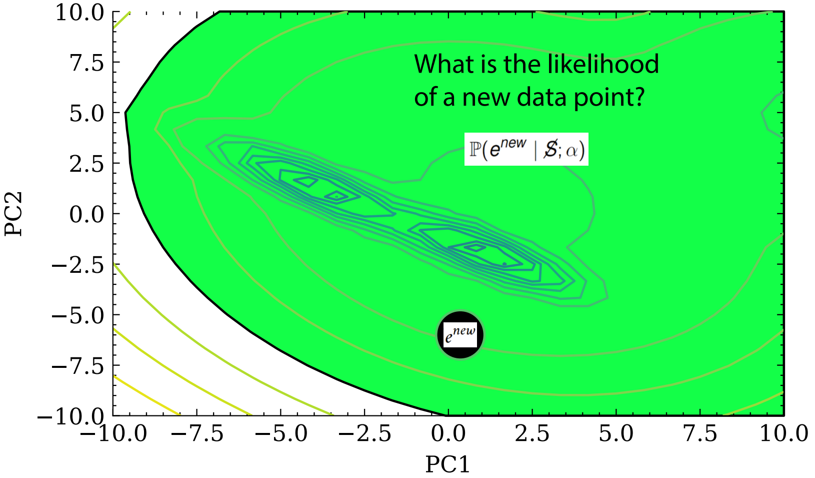
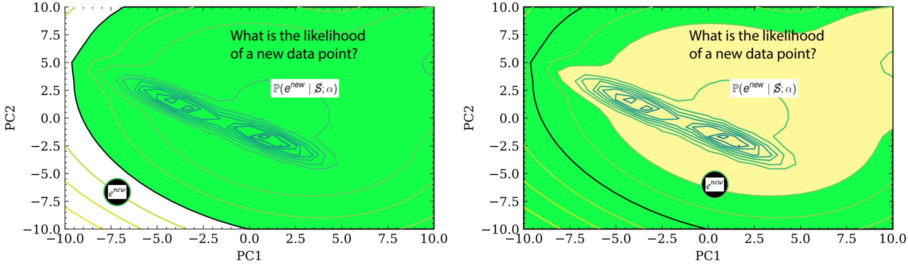


Figure 11: Illustration of $\not{S}$ with large α.

Inference Procedure



Inference Procedure



Computing Likelihood

$$\mathbb{P}(e^{new} \mid \not{S}; \alpha) = \begin{cases} 0 & \text{if } e^{new} \in S \\ \mathrm{Vol}(L_{e^{new}}\text{-LDR}) & \text{if } e^{new} \in \not{S} \end{cases}$$

Inference Procedure

Computing Marginal Likelihood

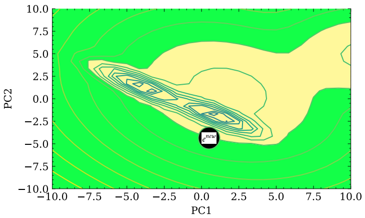
$$\mathbb{P}(e^{new}) = \mathrm{Vol}(L_{e^{new}}\text{-LDR})$$

Estimating w/ Monte Carlo

For a new sample:

$$\mathbb{P}(e^{new}) \leftarrow \frac{|\{e_i \mid \hat{p}(e_i) \leq \hat{p}(e^{new})\}|}{|\{e_i\}|}$$

where $\{e_i\}$ is the training set.



Priors

Non-informative prior

$$\mathbb{P}(\cancel{S_t}) = C$$

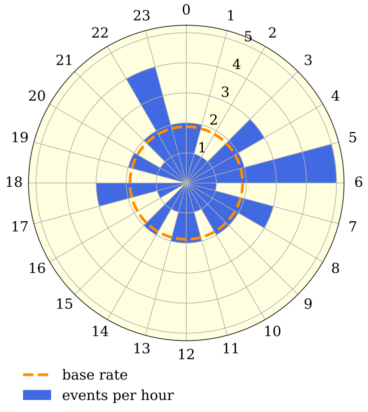
Informative prior

$$f(x|\mu, \kappa; \omega) = \frac{\exp(\kappa \cos(\omega(x - \mu)))}{2\pi I_0(\kappa)}$$

$$\mathbb{P}(S_t) = \frac{1}{K}\sum_{i=0}^{23} a_i f(t \mid i, k)$$

where a_i are the histogram values and K is calculated numerically.

Empirical Circadian Profile



Priors

Non-informative prior

$$\mathbb{P}(\cancel{S_t}) = C$$

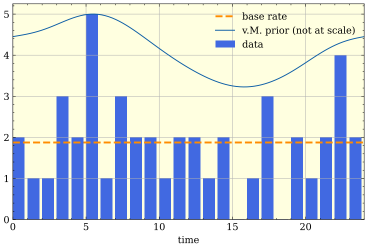
Informative prior

$$f(x|\mu,\kappa;\omega) = \frac{\exp(\kappa\cos(\omega(x-\mu)))}{2\pi I_0(\kappa)}$$

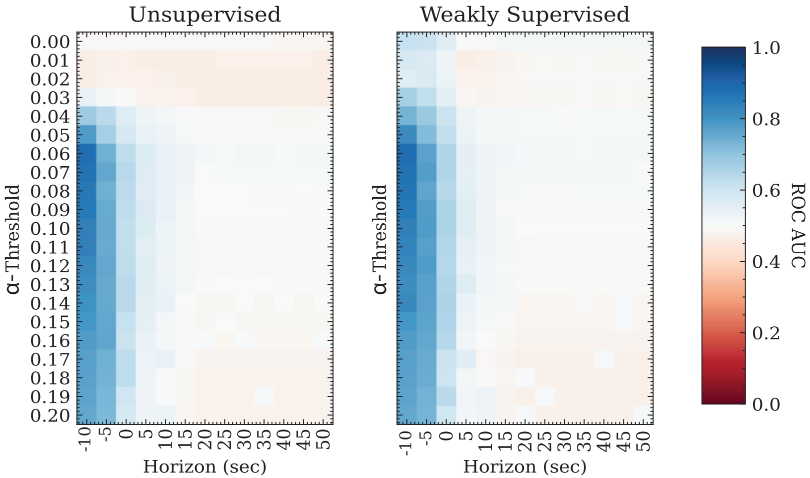
$$\mathbb{P}(S_t) = \frac{1}{K}\sum_{i=0}^{23} a_i f(t \mid i, k)$$

where a_i are the histogram values and K is calculated numerically.

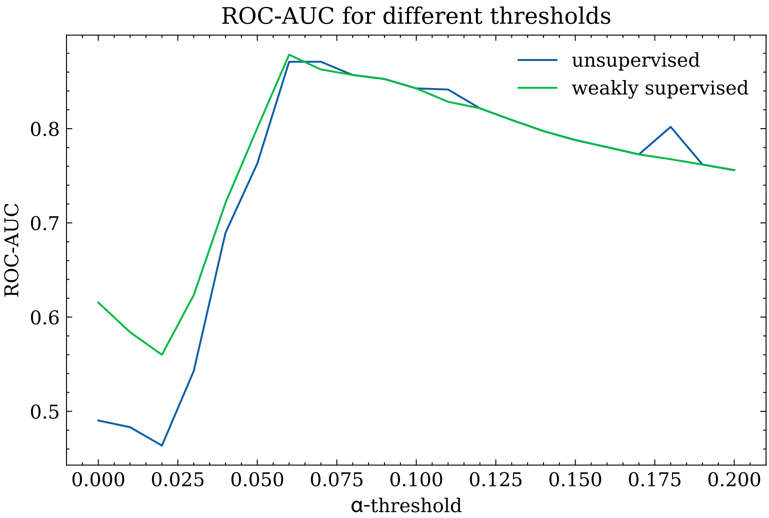
Empirical Circadian Profile



Evaluating BSLE on validation data



Evaluating BSLE on validation data



Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a seizure is not occurring, and a temporal prior to incorporate intentional bias from labels, if available.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a seizure is not occurring, and a temporal prior to incorporate intentional bias from labels, if available.

BSLE can be used for unsupervised seizure alerts, and in addition learn from untaped seizures.

Conclusion

We present BSLE - a Bayesian Seizure Likelihood Estimator.

BSLE uses a novelty score to model the likelihood that a seizure is not occurring, and a temporal prior to incorporate intentional bias from labels, if available.

BSLE can be used for unsupervised seizure alerts, and in addition learn from untaped seizures.

More work is required to statistically validate findings. Further work could use BSLE as a pseudo-labeler module in self-supervised learning.

End

Thank You

References I

L. D. Coles, E. E. Patterson, W. D. Sheffield, J. Mavoori, J. Higgins, B. Michael, K. Leyde, J. C. Cloyd, B. Litt, C. Vite, et al. Feasibility study of a caregiver seizure alert system in canine epilepsy. *Epilepsy research*, 106(3):456–460, 2013.

A. B. Gardner, A. M. Krieger, G. Vachtsevanos, B. Litt, and L. P. Kaelbing. One-class novelty detection for seizure analysis from intracranial eeg. *Journal of Machine Learning Research*, 7(6), 2006.

R. J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996.

References II

P. J. Karoly, H. Ung, D. B. Grayden, L. Kuhlmann, K. Leyde, M. J. Cook, and D. R. Freestone. The circadian profile of epilepsy improves seizure forecasting. *Brain*, 140(8): 2169–2182, 2017.

Y. Yang, N. D. Truong, J. K. Eshraghian, A. Nikpour, and O. Kavehei. Weak self-supervised learning for seizure forecasting: a feasibility study. *Royal Society Open Science*, 9(8):220374, 2022.

Apx. 1 - Gaussian processes

Definition (Gaussian process)

$$f(t) \sim \mathcal{GP}(m(t), k(t, t'))$$

where

$$m(t) = \mathbb{E}[f(t)]$$

$$k(t, t') = \mathbb{E}[(f(t) - m(t))(f(t') - m(t'))]$$

Apx. 1 - Gaussian processes

Definition (Gaussian process)

$$f(t) \sim \mathcal{GP}(m(t), k(t, t'))$$

where

$$m(t) = \mathbb{E}[f(t)]$$

$$k(t, t') = \mathbb{E}[(f(t) - m(t))(f(t') - m(t'))]$$

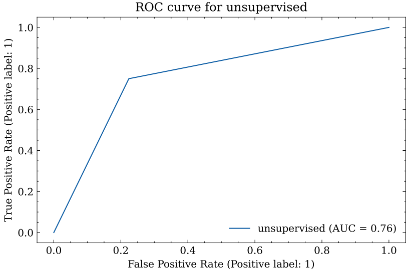
Definition (Matérn class of covariance functions)

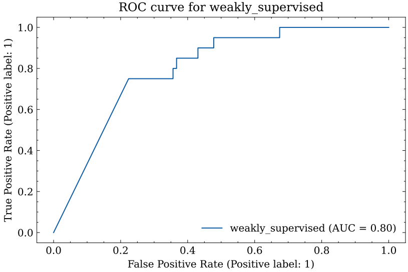
$$k_{Matern}(t, t') := \frac{2^{1-\nu}}{\Gamma(\nu)}(\sqrt{2\nu}d)^{\nu} K_{\nu}(\sqrt{2\nu}d)$$

where

$$d = (t - t')^T \Phi^{-2}(t - t')$$

Apx 2. - ROC curves





Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

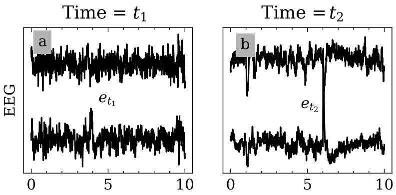


Figure 12: Samples from $E(t_1)$ and $E(t_2)$.

Sample spaces

Let:

1. $e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N}$

a random EEG variable,

and:

2. $s_t \sim S(t) \in \Omega_S = \{0, 1\}$

a random Seizure variable.

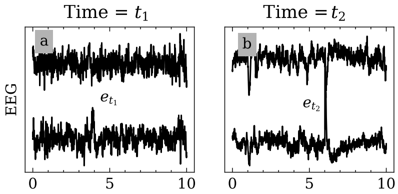


Figure 12: Samples from $E(t_1)$ and $E(t_2)$.

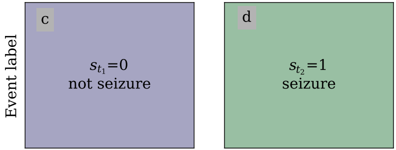


Figure 13: Samples from $S(t_1)$ and $S(t_2)$