Mathematical modelling and machine learning for medicine

Maria Veretennikova

University of Warwick, United Kingdom
Invited talk at the National Research University Higher School of Economics, Russia

1st of November 2020

Mathematical modelling and machine learning

- Using data for building a prediction/forecast model for future unseen cases
- Learning - parameter tuning for smooth and accurate decision making in the future
- Machine learning problems: Classification, regression, image reconstruction, anomaly detection, recommendation systems, time series change point detection
- fMRI/EEG signal cleaning

Tasks in medicine and mathematical techniques

- Healthy/ill patient, cancer type identification – decision trees, neural networks, support vector machines
- Optimal allocation of patients into treatments in a clinical trial – multi-armed bandits and other approaches
- Modelling drug diffusion through tissue, e.g. anti-nicotine drug patches on skin – diffusion models such as continuous time random walks and stochastic differential equations
- Modelling the optimal molecular drug structure – genetic algorithms, neural networks

Patient treatment choice as an optimization problem

Multi-armed bandit approach to treatment allocation has theoretically grounded strategies approximating the optimal one

Какое лекарство дать следующему пациенту?

Бандит 1 (группа - лекарство 1)	Бандит 2 (группа - лекарство 2)	Бандит 3 (группа - лекарство 3)
Успех 50%	**Успех 20%**	**Успех 70%**

SVM - patient classification

Support Vector Machines allow to find the equation of the optimally separating hyperplane for data points of different classes. The kernel trick mapping the data into another space.

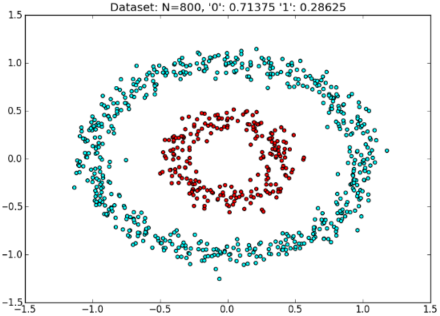


Image from: https://www.eric-kim.net/

SVM - patient classification

Computer-aided lung disease classification

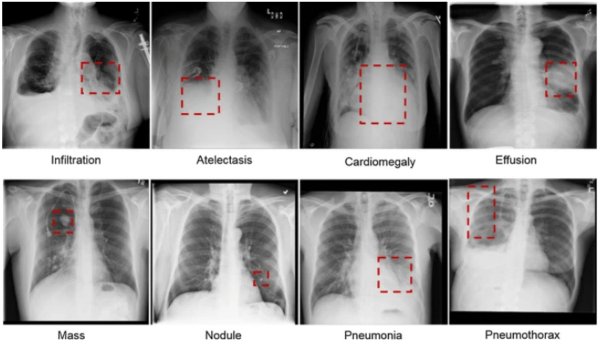


Image from: https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-018-0544-y

Data augmentation, synthetic sample generation

Generative Neural Networks may be employed for realistic synthetic image data generation in case there is a shortage of real data for drawing inference.

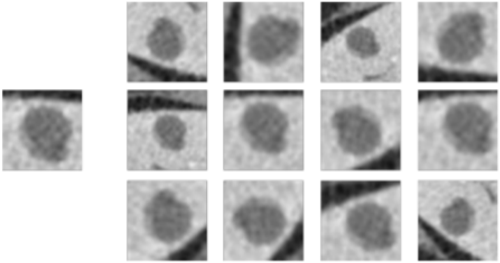
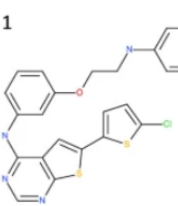


Image from: https://arxiv.org/pdf/1803.01229.pdf

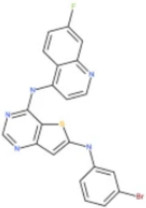
Molecular design for drugs

Mathematical tools include Genetic Algorithms and Generative Adversarial Networks

1



2



3

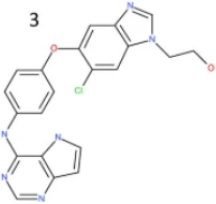
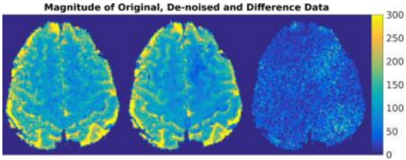
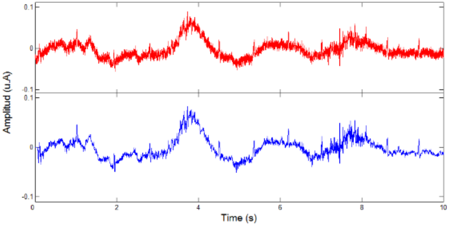


Image from:
https://jcheminf.biomedcentral.com/articles/10.1186/s13321-019-0397-9

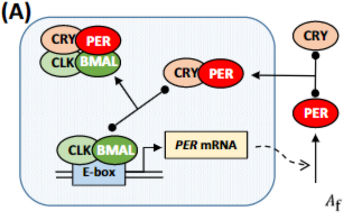
fMRI and EEG noise removal

In medical data noise could be due to electromagnetic interference, uncontrolled physiological processes affecting the signal, such as breathing, sudden muscle movements due to pain etc.





Gene expression oscillation mechanism may be modelled with a system of ODEs and mathematical tools allow to analyze characteristics of observed data for further inference



(B)

$$\frac{dM}{dt} = \beta\left(\frac{A_f}{A_T} - M\right)$$

$$\frac{dP_c}{dt} = \beta(M - P_c)$$

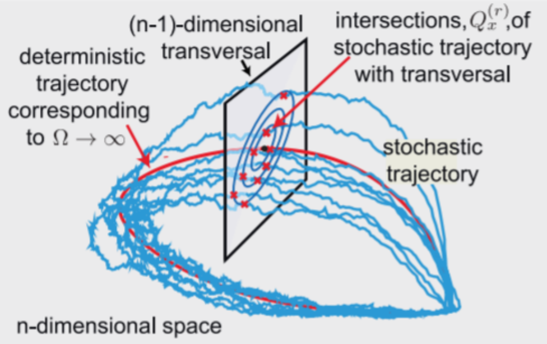
$$\frac{dP_n}{dt} = \beta(P_c - P_n)$$

$$A_f = \frac{1}{2}\left[A_T - P_n - K_d + \sqrt{(A_T - P_n - K_d)^2 + 4K_dA_T}\right]$$

Image from:
https://reader.elsevier.com/reader/sd/pii/S0962892420300738?token=73

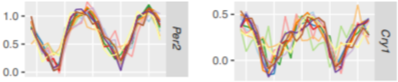
Simulation and analysis of data - pcLNA

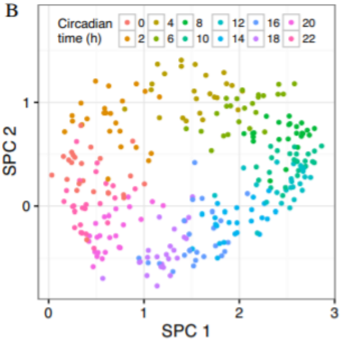
Generation of synthetic data for an oscillatory dynamics requires reliable techniques, which focus on realistic data distributions.



Here the red loop is the solution of an ODE system such as the one on the previous slide for gene expression oscillations. Mathematical modelling allows simulations for testing research ideas if real data is scarce.

ZeitZeiger is a machine learning tool developed to predict the timing gene expression data was obtained for accurate decision-making. This could be relevant for understanding the inner clock rhythm disruption for tailored medicine





EEG

Common EEG setup: 32 electrode channels, may be way more: all attached to the scalp of the head at specific points.

EEG allows to monitor electrical activity within the brain and measures the changes in electric potentials with a common reference to a quiet electrode

EEG uses include: locating brain damage, determining sleep stage, monitoring anesthesia depth

Frequency band separation with wavelets Db4

Splitting the signal into 7 frequency bands: Daubechies4 wavelets

Table: Frequency band correspondence

Traditional	Db4 band's central frequency
Delta 0 – 3.5 Hz	2.7 Hz
Theta 3.5 – 7.5 Hz	5.57 Hz
Alpha 7.5 – 13 Hz	11 Hz
Beta 13 – 30 Hz	22.3 Hz
Gamma $>$ 30 Hz	four subsequent bands

EEG and frequency bands

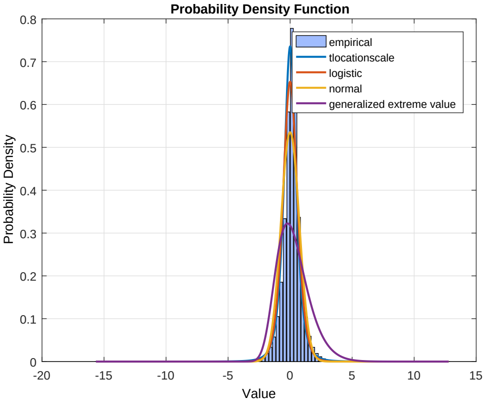
- Delta - Deep sleep, no dreams
- Theta - Daydreaming
- Alpha - Rest after mental activity
- Beta -Engaged mind
- Gamma - Intense brainstorming, brain disorders

Stochastic process approach to electroencephalogram (EEG) data

Our question: can we use stochastic processes to model EEG signals recorded during coma for different channels and use these to predict the neurodevelopmental level of patients 6 months after emergence from coma due to cerebral malaria?

- All frequency bands - increment process histograms: t-distribution, stationarity

Figure: Alpha frequency band



Stochastic processes

- Student Ornstein-Uhlenbeck process: suitable model, SDE

$$dX(t) = -\lambda X(t)dt + dY(\lambda t), t \geq 0, \tag{1}$$

  where $\{Y(t), t \geq 0\}$ is the background driving Levy process
- and the parameters of the t-distribution were used as features for the prediction problem

Other features and basic result

- Other EEG features: proportion of flat line in signal, frequencies of peaks higher than nearest neighbours by a certain proportion of signal's standard deviation, entropy and other standard time series characteristics
- Other non-EEG features: height, weight, age, numerous blood and cerebrospinal fluid characteristics
- The t-distribution parameters were better predictors as a group than medical/non-EEG based features, specific channels/brain regions identified as useful for further biomarker analysis

Methods used

- Data-preprocessing with Persyst: Neural networks for noise removal
- SVD-based matrix completion method for missing data: Soft Impute (Hastie, Tibshirani): an iterative matrix approximation technique
- Regularized regression: Lasso and Elastic Net regression used for the prediction problem, number of features $>>$ number of observations
- Leave-one-out cross validation used for choosing optimal tuning parameters

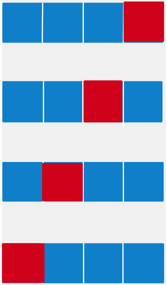
Methods - some details - regression and regularization

Let us denote by X the n by p normalized feature matrix, n=78, p - number of features considered, and y - the neurodev. level in 6 months. We solve the following optimization problem:

$$\min_{\bar{\beta}} \left(\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \alpha \times l_1 \sum_{j=1}^{p} |\beta_j| + 0.5 \times \alpha \times (1 - l_1) \sum_{j=1}^{p} \beta_j^2 \right)$$

and the tuning parameters l_1 and α are found by cross-validation, as explained on the next slide.

Methods - some details - cross-validation



1) Train on blue, test on red
2) Average error across blocks for each pair of hyperparameters

3) Select the pair giving the least average

Use for doctors

- Clinical trial design: prognostic factors to be used for patient allocation into different treatment groups - more research needed
- Determining patients with the worst prognosis for intensive care treatment

References

- Vapnik, V., Guyon, I. and Hastie, T., 1995. Support vector machines. Mach. Learn, 20(3), pp.273-297.
- Hughey, J.J., Hastie, T. and Butte, A.J., 2016. ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. Nucleic acids research, 44(8), pp.e80-e80.
- Veretennikova, M.A., Sikorskii, A. and Boivin, M.J., 2018. Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria. Journal of statistical distributions and applications, 5(1), pp.1-12.
- Minas, G. and Rand, D.A., 2017. Long-time analytic approximation of large stochastic oscillators: Simulation, analysis and inference. PLoS computational biology, 13(7), p.e1005676.

References

- Villar, S.S., Bowden, J. and Wason, J., 2015. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. Statistical science: a review journal of the Institute of Mathematical Statistics, 30(2), p.199.
- Relógio, A., Westermark, P.O., Wallach, T., Schellenberg, K., Kramer, A. and Herzel, H., 2011. Tuning the mammalian circadian clock: robust synergy of two loops. PLoS Comput Biol, 7(12), p.e1002309.
- fMRI images from: https://www.intechopen.com/books/adaptive-filtering-applications/noise-removal-from-eeg-signals-in-polisomnographic-records-applying-adaptive-filters-in-cascade
- EEG signal image from: http://indexsmart.mirasmart.com/ISMRM2017/PDFfiles/1462.html