AIMLab.

ML IN HEALTHCARE
FACULTY OF BIOMEDICAL ENGINEERING

Contents

1. Introduction

With billions of mobile devices worldwide and the low cost of connected medical sensors, recording and transmitting medical data has become easier than ever. However, this ‘wealth’ of physiological data has not yet been harnessed to provide actionable clinical information. This is due to the lack of smart algorithms that can exploit the information encrypted within these ‘big databases’ of biomedical time series and images, take individual variability into account and generalize to different population sample.

Exploiting such data necessitates an in depth understanding of the physiology underlying the biomedical time series and images, the use of advanced digital signal processing and machine learning tools (e.g. deep learning) to recognize and extract characteristic patterns of health function, and the ability to translate these patterns into clinically actionable information for the purpose of **diagnosis, prognosis and treatment**. In particular, the creation of intelligent algorithms combined with existing and novel wearable and biosensors offer an unprecedented opportunity to improve **Human Health** by providing new intelligent patient monitoring systems in the clinical environment and for remote health monitoring.

In this course you will learn about aspects of information processing including data preprocessing, visualization, regression, dimensionality reduction (PCA, ICA), feature selection, classification (LR, SVM, Deep Learning) and their usage for decision support in the context of **biomedical engineering** and with a focus on improving **Human Health**. It will aim to train a new generation of scientists whom can perform research on large steams of data including **genomic data, sensor data and healthcare data.** The course aims to provide an overview of computer tools and machine learning techniques for processing such datasets within the context of healthcare. Each session is structured with two lectures and two hours of tutorial plus an optional third hour of “workshop”. During the lectures the necessary theory and intuition will be covered and practical (“hands on”) computer based tutorials and assignments will confront you with real world research question dealing with a variety of medical datasets. The lectures are divided in three parts: ML basis, popular classifiers and introduction to deep learning.

2. Course summary

Course title:	Machine Learning in Healthcare (MLH)
Short title:	ML in Healthcare
Course ref. no.	336546
Number of credits:	3
Number of weeks: - **Weekly lectures** - **Weekly tutorials**	13 2 hours (total 26 hours) 2 hours (total 26 hours) + 1 hour optional (13 hours)
Course assessment:	Four assignments: 20%-30%-15%-20%. Final oral exam: 15%. Attendance for both lectures and tutorials is mandatory for at least 70% of meetings, i.e. you cannot miss more than four lectures and four tutorials.
Capacity:	32 Working station
Computer requirements:	Six GPU (department cluster). Software: PyCharm, jupyter notebook, Git, Atom. Libraries: Numpy, Panda, Scikit-learn, Keras.
Lecturer(s):	Joachim A. Behar (JB), PhD
Teaching assistants:	Jonathan Fhima (JF), PhD candidate Yevgeniy Men (YM), MSc candidate Anat Rotschield (AR), MSc candidate
Guests Lecturers:	Anne Weill (AW), PhD, Technion-BME Danny Eytan (DE), MD-PhD, Rambam Hospital
Teaching objectives:	Students will acquire the following skills: • Python for biomedical data science. • Main classifiers, intuition and mathematical background. • Neural networks and deep learning. • Performance statistics in healthcare. • ML for diagnosis, prognosis and treatment. • Ground truth in medical data science.

3. Syllabus

1.1 Part I: ML Basis

W	Lecture	Subjects covered
1	BME-336546-L01-Introduction to machine learning in healthcare	- Course objectives and settings - Introduction to ML in healthcare - Supervised, unsupervised and reinforcement learning - ML for diagnosis, prognosis and treatment - Medical data, sources, challenges and regulations - Polynomial curve fitting - Cost function - Under and overfitting - Notations
2	BME-336546-L02-Data exploration and preprocessing	- Exploratory data analysis - Data visualization - Abnormality detection and handling - Features scaling
	BME-336546-L03-Linear models for regression	- Intuition - Calculus proof - Probabilistic proof - Sequential learning - Cost function
3	BME-336546-L04-Linear models for classification	- Classification versus regression - LR hypothesis representation - Cross entropy - Gradient descent - Multiclass classification: one against all, multinomial
	BME-336546-L05-Odds and odds ratio	- Odds ratio - Confounding
4	BME-336546-L06-Regularization	- Overfitting - Cost function - Regularized linear regression - Regularized logistic regression - Ridge, Lasso regression - Geometrical interpretation
	BME-336546-L07-Practical consideration on training a model	- Evaluating a model: train, validation and test sets - Model selection, learning curves and error analysis - Bias-variance tradeoff - Cross validation approaches - Stratification - Information leakage - Generalization performance
5	BME-336546-L08-Performance statistics	- Performance statistics - Receiver operative curve - Multiclass classification - Training the final ML model
	Guest speaker Rambam	- Application of ML in clinical practice

1.2 Part II: Popular classifiers

W	Lecture	Subjects covered
6	BME-336546-L09-Introduction to nonlinear models	- Linear but with nonlinear features - Change of basis
	BME-336546-L10-Support vector machines	- Maximum margin classifiers - Dual representation - Kernel trick - Grid search and random search
7	BME-336546-L11-Feature selection	- Relevance and redundancy - Filters, wrappers and embedded - LASSO, mRMR
8	BME-336546-L12-Unsupervised learning with k-means and Gaussian mixture models	- K-nearest neighbor - Probabilistic data analysis: GMM
9	BME-336546-L13-Principal component analysis	- Blind source separation - Principal component analysis - Change of basis - Mathematical proof - PCA in machine learning
	BME-336546-L14-Independant component analysis	- Independent component analysis - Statistical independence versus correlation - Whitening - Beyond ICA: t-SNE

1.3 Part III: Neural networks and introduction to deep learning

Week	Lecture	Subjects covered
10	#C21 ANN I: introduction	- Revisiting logistic regression - Introduction to NN - Notations - Representation learning - Forward propagation - Backward propagation - Activation functions - Multiclass classification (softmax)
	#C22 ANN II: training a NN	- Revisiting train-validation-test split - Weight initialization - Optimization algorithms - Revisiting bias-variance tradeoff - Batch normalization
11	#GL01 Performance computing	- High Performance Computing (AW)
	#GL01 Performance computing	- High Performance Computing (AW)
12	#C23 ANN III: hyperparameters tuning	- Grid search - Random search - Bayesian optimization - Vanishing and exploding gradient
	#C24 Deep Learning CNN	- Foundation - Convolution - CNN architecture - Striding, padding, pooling
13	#C25 Deep Learning CNN	- CNN architectures
	#C26 Examples of medical ML @Technion	- Presentation of ongoing research in the lab.