Deep models for temporal data with applications to electrocardiography

Daniel Gedon

Department of Information Technology,
Uppsala University.

Half-time seminar
March 25, 2022

Outline

UPPSALA UNIVERSITET

Outline

UPPSALA UNIVERSITET

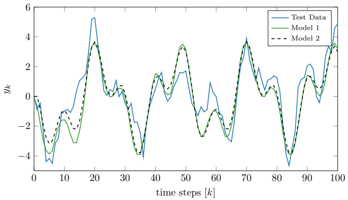
UPPSALA UNIVERSITET

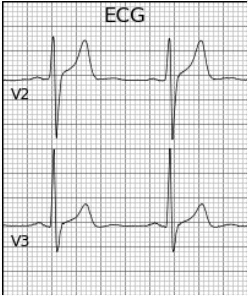
Deep models for temporal data with applications to electrocardiography

Success of deep learning based on:

- powerful computers
- large data sets

$\Rightarrow$ Many natural data sets are collected over time





In this talk: Different types of modeling temporal data

Part 1:

[Paper I] Deep State Space Models for Nonlinear System Identification
D. Gedon, *N. Wahlström, T. Schön, L. Ljung*
19th IFAC Symposium on System Identification (SYSID), 2021

Part 2:

[Paper II] Automatic 12-lead ECG Classification Using a Convolutional Network Ensemble
A.H. Ribeiro, **D. Gedon**, *D. Teixeira, M.H. Ribeiro, A.L. Pinho Ribeiro, T. Schön, W. Meira Jr.*
Computing in Cardiology (CinC), 2020

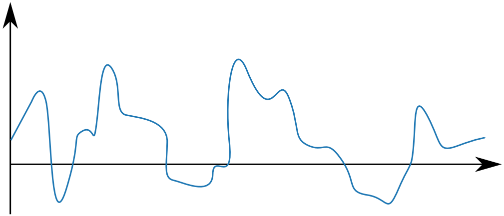
[Paper III] First Steps Towards Self-Supervised Pretraining of the 12-Lead ECG
D. Gedon, *A.H. Ribeiro, N. Wahlström, T. Schön*
Computing in Cardiology (CinC), 2021

[Paper IV] Artificial Intelligence-Based ECG Diagnosis of Myocardial Infarction in Emergency Department Patients
*S. Gustafsson**, **D. Gedon***, *E. Lampa, A.H. Ribeiro, M. Holzmann, T. Schön, J. Sundström*
NeurIPS Workshop, 2021
Submitted to Circulation, 2022

+ Ongoing work

Outline

Modeling with temporal data



Dynamical system view

- time-series to time-series
- Model: replicate system dynamics

Prediction view

- time-series to point
- Model: Classifier / Regressor

Outline

UPPSALA UNIVERSITET

- Dynamical system view = System identification
- From input-output data to one-step ahead predictor $f_\phi(\cdot)$

$$\{(\mathbf{u}_t, \mathbf{y}_t)\}_{t=1}^{T} \quad \Rightarrow \quad \hat{\mathbf{y}}_{t+1} = f_\phi(\mathbf{u}_{1:t}, \mathbf{y}_{1:t})$$

One way to achieve this:

[Paper I] Deep State Space Models for Nonlinear System Identification

- SSM given as

$$\mathbf{h}_t = f_\theta(\mathbf{h}_{t-1}, \mathbf{u}_t, \mathbf{y}_t).$$
$$\hat{\mathbf{y}}_{t+1} = g_\theta(\mathbf{h}_t).$$

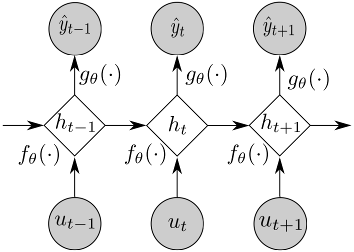
- deep SSM as extension of classic SSM with Neural Networks
- show that deep SSM are useful for system identification
- Pedagogical paper: explain deep SSM to system identification community

Dynamical system view - Deep SSM

UPPSALA UNIVERSITET

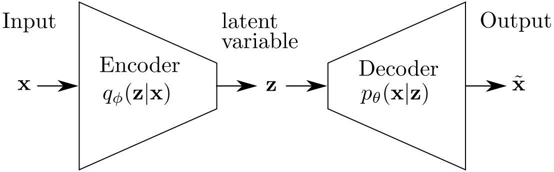
Recurrent Neural Network (RNN)

- Recursive propagation of hidden state.
- Dirac delta function as state transition distribution.



Variational Autoencoder (VAE)

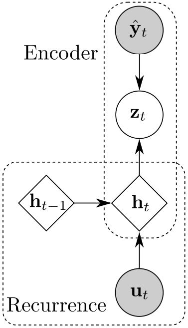
- Decoder: $p_\theta(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}^{dec}, \boldsymbol{\sigma}^{dec})$,
$[\boldsymbol{\mu}^{\mathrm{dec}}, \boldsymbol{\sigma}^{\mathrm{dec}}] = \mathrm{NN}_\theta^{\mathrm{dec}}(\mathbf{z})$.
- Prior: $p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$.
- Encoder: $q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}^{\mathrm{enc}}, \boldsymbol{\sigma}^{\mathrm{enc}}\right)$,
$[\boldsymbol{\mu}^{\mathrm{enc}}, \boldsymbol{\sigma}^{\mathrm{enc}}] = \mathrm{NN}_\phi^{\mathrm{enc}}(\mathbf{x})$.



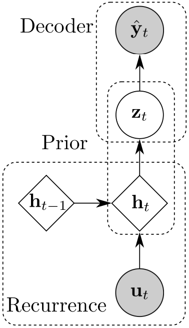
Dynamical system view - Deep SSM

- Require a temporal extension of the VAE.
  $\rightarrow$ combine RNN and VAE
- Prior: update with RNN output,
  $p_\theta(\mathbf{z}_t|\mathbf{h}_t) = \mathcal{N}\left(\mathbf{z}_t|\boldsymbol{\mu}_t^{\text{prior}}, \boldsymbol{\sigma}_t^{\text{prior}}\right),$
  $[\boldsymbol{\mu}_t^{\text{prior}}, \boldsymbol{\sigma}_t^{\text{prior}}] = \text{NN}_\theta^{\text{prior}}(\mathbf{h}_t).$

[Paper I] studies six variants of deep SSM



Inference network



Generative network

Dynamical system view - Deep SSM results

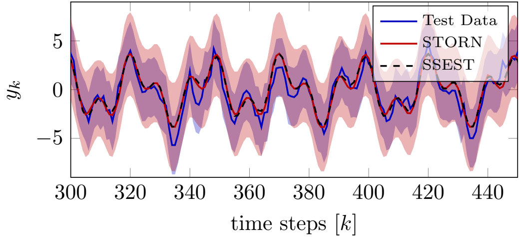
Linear Gaussian system:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.7 & 0.8 \\ 0 & 0.1 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} -1 \\ 0.1 \end{bmatrix} \mathbf{u}_k + \mathbf{v}_k,$$

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k,$$

$$\mathbf{v}_k \sim \mathcal{N}(0, 0.5 \cdot \mathbf{I}), \ \mathbf{w}_k \sim \mathcal{N}(0, 1).$$

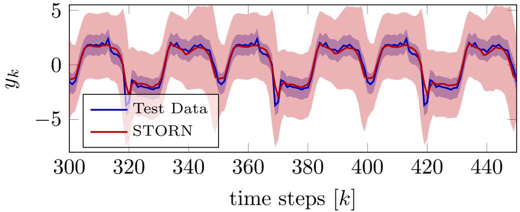
Toy Problem: Linear Gaussian System



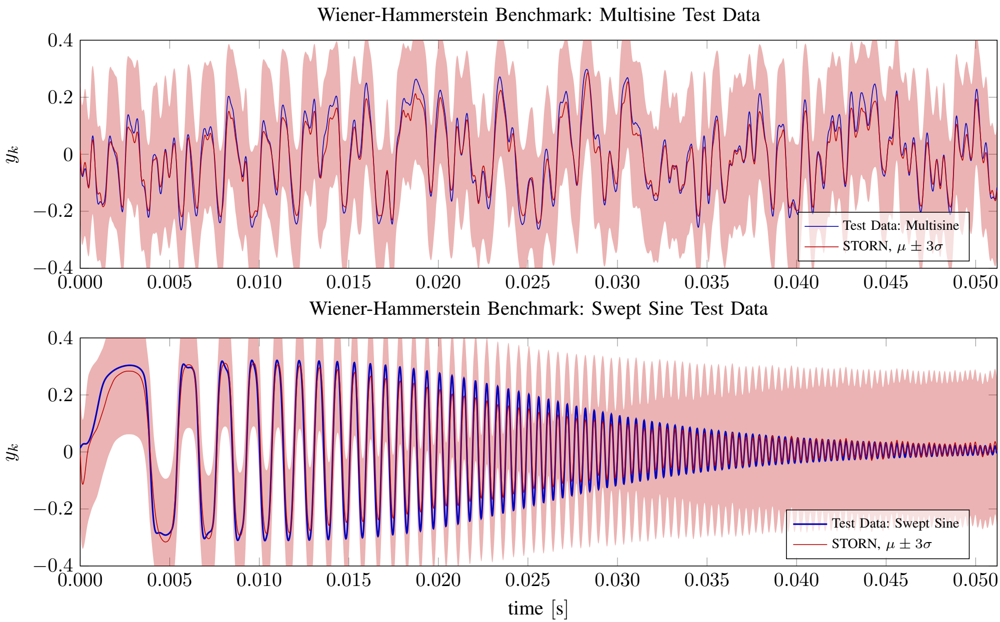
Nonlinear Narendra-Li benchmark:

- highly nonlinear dynamics
- two states

Narendra-Li Benchmark



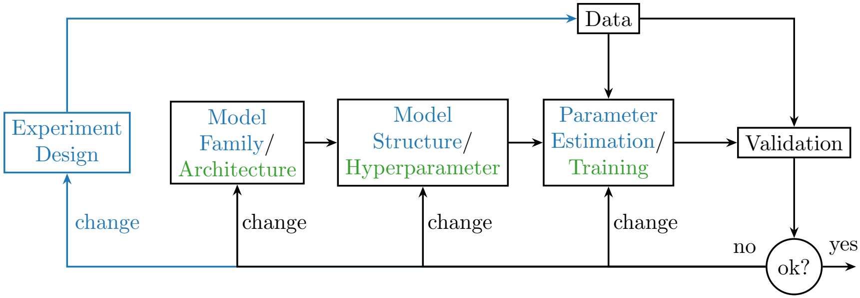
Dynamical system view - Deep SSM results



UPPSALA UNIVERSITET

Ongoing work: Deep Learning for System Identification Survey

- Deep SSM is just one method
- Lots of overlap between system identification and deep learning method
- Example: general modeling procedure



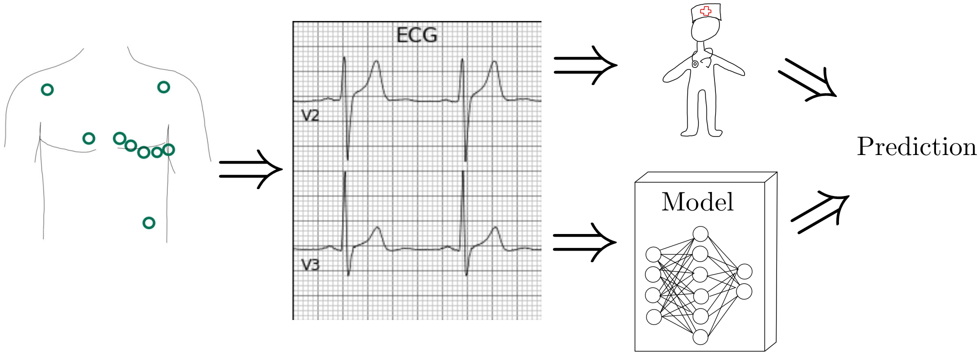
Outline

Prediction view - General

Prediction view = obtain a point prediction

$$\mathbb{R}^t \to \mathbb{R} \quad \text{for regression}$$

$$\mathbb{R}^t \to \mathbb{Z} \quad \text{for classification}$$



Prediction view - General

Problems we are facing:

- Data of varying length

[Paper II] Automatic 12-lead ECG Classification Using a Convolutional Network Ensemble

- Lots of unlabeled data

[Paper III] First Steps Towards Self-Supervised Pretraining of the 12-Lead ECG

- Model improvements

[Paper IV] Artificial Intelligence-Based ECG Diagnosis of Myocardial Infarction in Emergency Department Patients

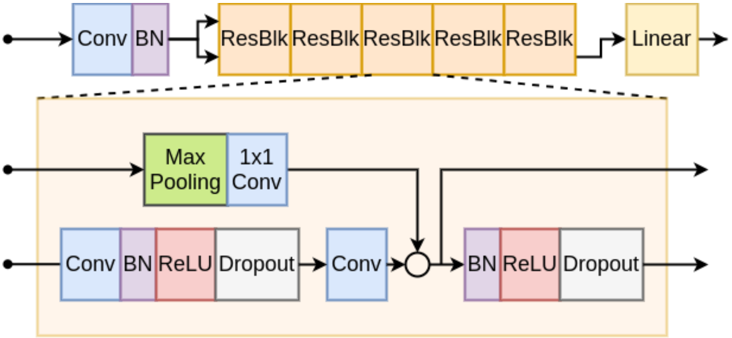
- Label noise
- Uncertain predictions

Ongoing work

Prediction view - Varying length

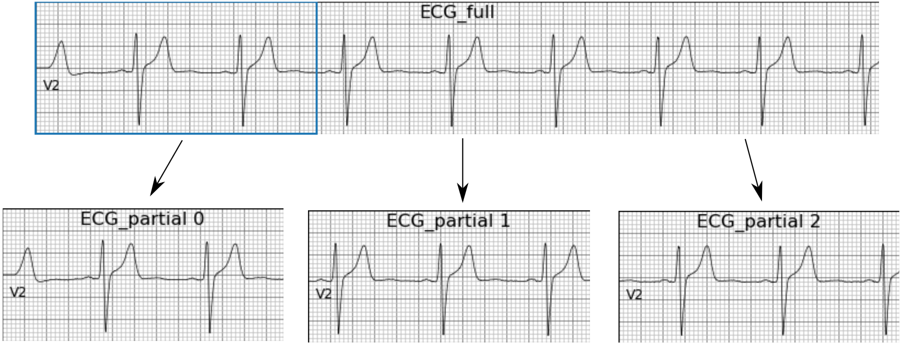
[Paper II] Automatic 12-lead ECG Classification Using a Convolutional Network Ensemble

Model of choice: adapted ResNet

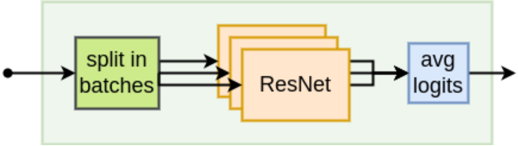


$\Rightarrow$ Requires fixed input size

How to deal with varying input size?

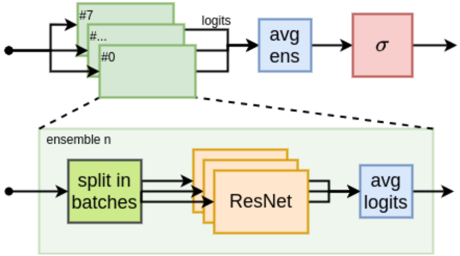


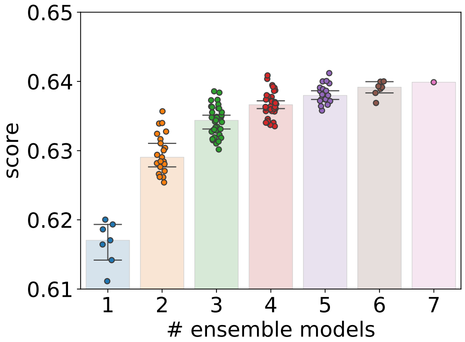
How to combine predictions?



UPPSALA UNIVERSITET

Can we use this method to improve overall performance?





Prediction view - Lots of unlabeled data

UPPSALA UNIVERSITET

[Paper III] First Steps Towards Self-Supervised Pretraining of the 12-Lead ECG

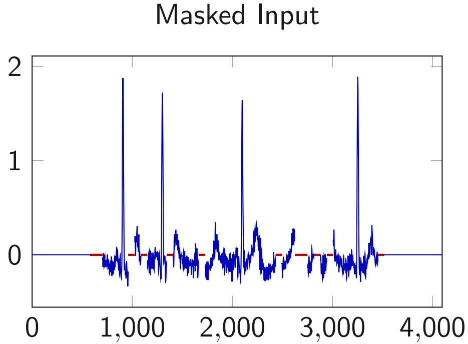
- Most data is collected without high quality labels
- Raw data itself contains lots of information
- How to utilize this information?

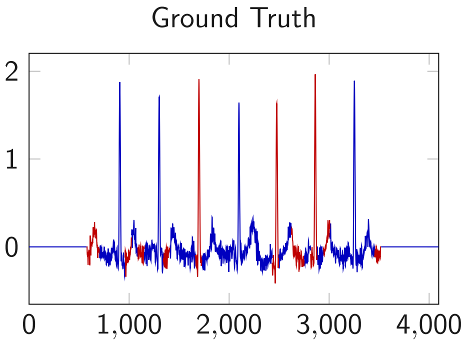
$\Rightarrow$ Use self-supervised methods

- Generate label from the signal itself

1. Self-supervised pre-training (lots of data)
2. Fine tuning on downstream task (limited data)

- Completion based generation of labels.
- Input: Replace subsequences of chosen length with zero.
- Output: Predict masked subsequences.

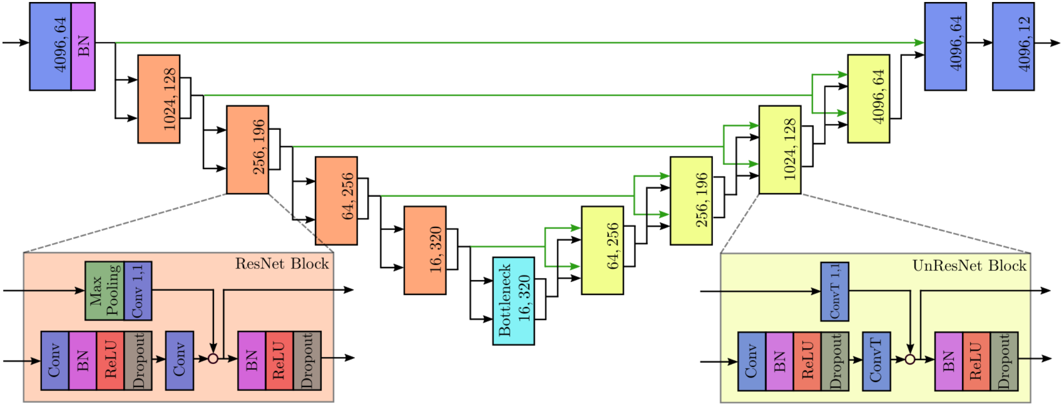




- Problem: How many samples to mask? Too easy vs too hard completion.

Model architecture

U-ResNet: ResNet based encoder-decoder + U-Net skip connections.



Prediction view - Model improvements

[Paper IV] Artificial Intelligence-Based ECG Diagnosis of Myocardial Infarction in Emergency Department Patients

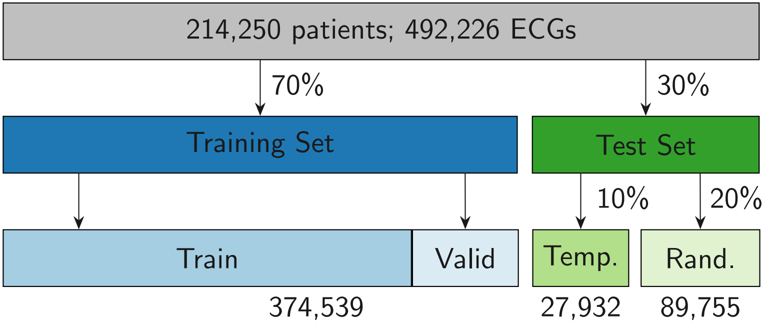
⇒ Solve difficult real world problem.

- Myocardial Infarctions (MIs):
  - 9M deaths/year, 200M disability-adjusted life years/year, and rising.
  - False negatives: 10-50,000 missed cases/year at EDs in the United States.
  - False positives: Less than half of those hospitalized for a suspected MI are diagnosed.

    → High burden on public health.
- ECG:
  - ST-elevation MI (STEMI) → detect in ECG
  - non-ST-elevation-MI (NSTEMI) → require blood testing

Prediction view - Model improvements

Splitting of the data set:

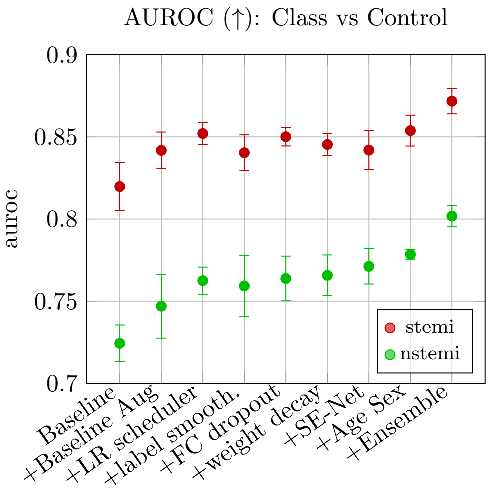


Highly unbalanced data set:

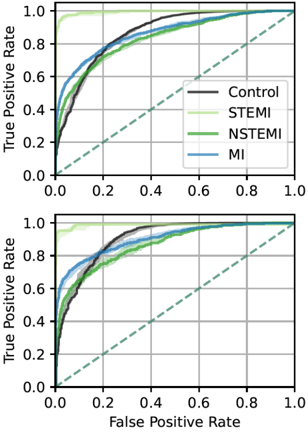
	Control	STEMI	NSTEMI
Absolute	484,992	1,818	5,416
Relative	98.5%	0.4%	1.1%

Prediction view - Model improvements

Results from model improvement on smaller data set



Final results



ROC curve.

Top: temporal split.

Bottom: random split.

Numeric results on test sets

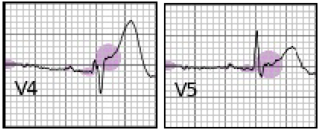
		Random	Temporal
AUROC (↑)	Control	0.86 (0.004)	0.90 (0.005)
	STEMI	0.99 (0.002)	0.99 (0.001)
	NSTEMI	0.83 (0.004)	0.87 (0.006)
	MI	0.86 (0.004)	0.90 (0.004)
Brier (↓)	Control	0.18 (0.000)	0.18 (0.000)
	STEMI	0.05 (0.000)	0.05 (0.000)
	NSTEMI	0.05 (0.000)	0.05 (0.000)

Prediction view - Model improvements

Model analysis

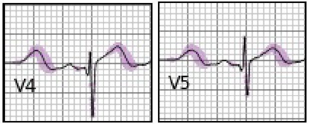
Grad-CAM plots $\rightarrow$ identify patterns of the model

STEMI



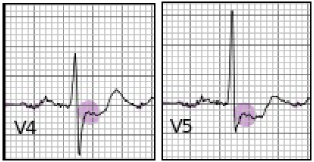
- ST-segment elevation
- typical for humans

STEMI



- Down-sloping T-wave
- untypical for humans

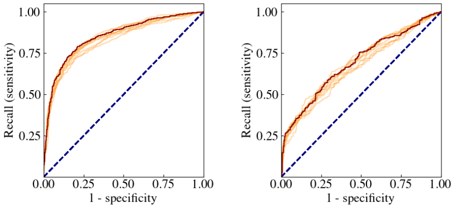
NSTEMI



- ST-segment depression
- humans would not suspect a MI

Further real-world data problems:

1. Label noise
   - Chagas prediction
   - Self-reported noisy labels
   - Mismatch of training and test label quality
2. Uncertain predictions
   - Predict electrolyte values
   - Regression
   - Use Gaussian/Laplace approximations



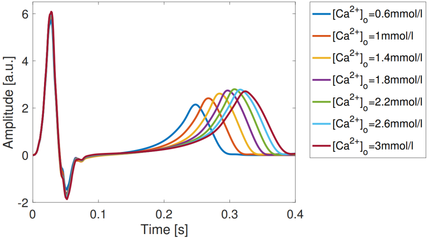
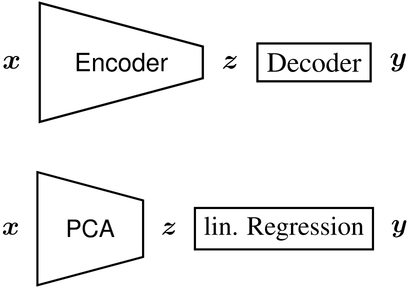


Figure from N. Pilia et al. "ECG as a tool to estimate potassium and calcium concentrations in the extracellular space," CinC, 2017

Outlook - Ongoing work

Inspired by self-supervised models:

- Self-supervised as specific unsupervised model
- Try to understand properties of unsupervised models
- Specifically analyse simple overparametrized models



Contact

Daniel Gedon, Uppsala University

E-mail: `daniel.gedon@it.uu.se`
GitHub: `https://github.com/dgedon`
Twitter: `@danigedon`

Supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by Knut and Alice Wallenberg Foundation.