Deep latent variable models for longitudinal biomedical data

Harri Lähdesmäki

Department of Computer Science
Aalto University

March 20, 2023

Research group: Computational systems biology

Research topics:

1. Probabilistic machine learning
   - Deep generative models
   - neural ODEs/PDEs/SDEs, GP ODEs
   - Gaussian processes
   - Longitudinal data analysis
2. Computational biology and biomedicine
   - Immune cell receptor sequence analysis
   - Sequencing data analysis
   - Biological network modeling
3. Applications in molecular biology and biomedicine

Research group:

- 2 postdocs
- 8 PhD students
- 3 research assistants

`https://research.cs.aalto.fi/csb/`

`harri.lahdesmaki@aalto.fi`

What is longitudinal data?

Time-series data that consist of

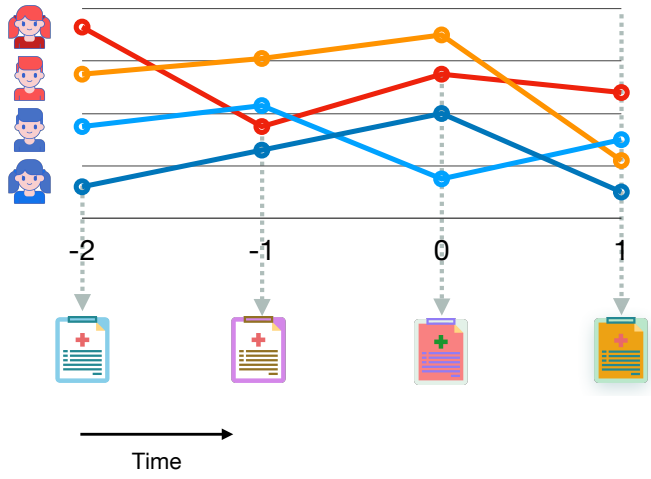
- Multiple subjects
- Each subject measured repeatedly over time

Observations have correlations

- Within a subject
- Across multiple subjects

Longitudinal data typically have

- High number of subjects
- Low number of measurements per subject



What models exist?

Generalized linear mixed models (GLMM)

$$\mathbf{y} = \underbrace{X\boldsymbol{\beta}}_{\text{fixed}} + \underbrace{Z\boldsymbol{\gamma}}_{\text{random}} + \underbrace{\boldsymbol{\epsilon}}_{\text{random}},$$

where $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 I)$

Interpretable, fast, powerful, good software support

What models exist?

Generalized linear mixed models (GLMM)

$$\mathbf{y} = \underbrace{X\boldsymbol{\beta}}_{\text{fixed}} + \underbrace{Z\boldsymbol{\gamma}}_{\text{random}} + \underbrace{\boldsymbol{\epsilon}}_{\text{random}},$$

where $\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 I)$

Interpretable, fast, powerful, good software support

Several other models:

- GAMs, local polynomials, splines, hierarchical Bayesian models, Gaussian processes, Gaussian process ANOVA, etc.

LonGP and *lgpr* models

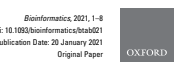
ARTICLE

https://doi.org/10.1038/s41467-019-09785-8 OPEN

An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data

Lu Cheng[1,2], Siddharth Ramchandran[1], Tommi Vatanen[3,4], Niina Lietzén[5], Riitta Lahesmaa[5], Aki Vehtari[1] & Harri Lähdesmäki[1]

Bioinformatics, 2021, 1–8
doi: 10.1093/bioinformatics/btab021
Advance Access Publication Date: 20 January 2021
Original Paper



Genetics and population analysis

***lgpr:* an interpretable non-parametric method for inferring covariate effects from longitudinal data**

Juho Timonen *, Henrik Mannerström, Aki Vehtari and Harri Lähdesmäki*

Additive Gaussian processes for longitudinal data

For several existing longitudinal models:

1. The model for the unknown function is additive:

$$f(\mathbf{x}) = f^{(1)}(\mathbf{x}) + \ldots + f^{(J)}(\mathbf{x})$$

2. Each $f^{(j)}(\mathbf{x})$ depends only on a small subset of variables

Additive Gaussian processes for longitudinal data

For several existing longitudinal models:

1. The model for the unknown function is additive:

$$f(\mathbf{x}) = f^{(1)}(\mathbf{x}) + \ldots + f^{(J)}(\mathbf{x})$$

2. Each $f^{(j)}(\mathbf{x})$ depends only on a small subset of variables

We assume that each additive component has an independent Gaussian process (GP) prior

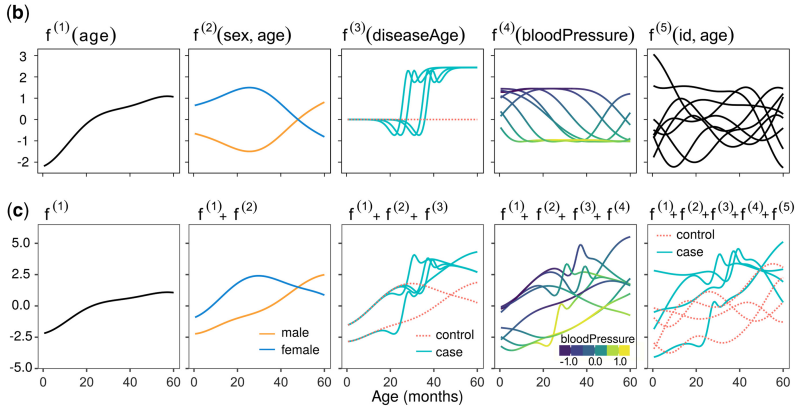
$$f^{(j)}(\mathbf{x}) \sim GP(0, \alpha_j^2 k_j(\mathbf{x}, \mathbf{x}' \mid \theta^{(j)})),$$

thus

$$f(\mathbf{x}) \sim GP\left(0, \sum_{j=1}^{J} \alpha_j^2 k_j(\mathbf{x}, \mathbf{x}' \mid \theta^{(j)})\right)$$

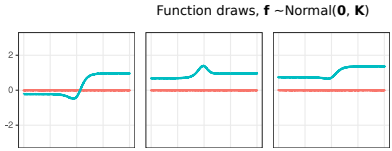
Kernel functions for longitudinal data

- Shared effects: the squared exponential (SE) kernel
- Category effects: product of the zero-sum and SE kernels
- Nonstationary shared effects: SE kernel with monotonic nonlinear input warping
- Individual random effects



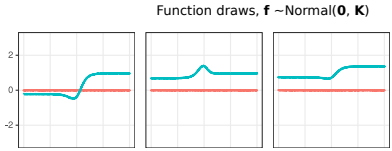
Kernel functions for disease effect

A. Non-stationary kernel $k_{\text{ns}}(\cdot)$ for the diseased individuals



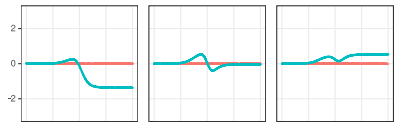
Kernel functions for disease effect

A. Non-stationary kernel $k_{\text{ns}}(\cdot)$ for the diseased individuals



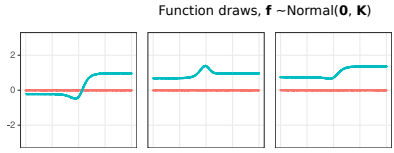
B. Variance masking: zero variance prior to disease onset

$$k_{\text{vm}}(x, x' \mid a, \ell) = f^a_{\text{vm}}(x) \cdot f^a_{\text{vm}}(x') \cdot k_{\text{ns}}(x, x' \mid a, \ell)$$



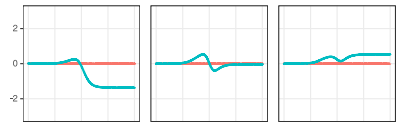
Kernel functions for disease effect

A. Non-stationary kernel $k_{\text{ns}}(\cdot)$ for the diseased individuals



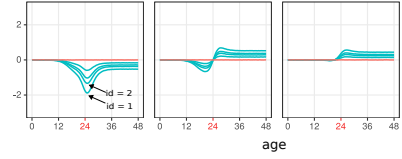
B. Variance masking: zero variance prior to disease onset

$$k_{\text{vm}}(x, x' \mid a, \ell) = f^a_{\text{vm}}(x) \cdot f^a_{\text{vm}}(x') \cdot k_{\text{ns}}(x, x' \mid a, \ell)$$



C. Heteregeneous vm kernel: different disease effect magnitude $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_Q]$ for each case individual

$$\sqrt{\beta_q \beta_{q'}} \cdot k_{\text{vm}}(x_{\text{disAge}}, x'_{\text{disAge}} \mid a, \ell_{\text{disAge}})$$



Bayesian inference and covariate relevance

Robust prior specifications

- $\boldsymbol{\theta}_{\text{kernel}}$: all kernel (hyper)parameters
- $\boldsymbol{\theta}_{\text{obs}}$: params. of the observation model
- $\boldsymbol{\theta}_{\text{other}}$: other parameters (input uncertainty etc.)

Bayesian inference and covariate relevance

Robust prior specifications

- $\boldsymbol{\theta}_{\text{kernel}}$: all kernel (hyper)parameters
- $\boldsymbol{\theta}_{\text{obs}}$: params. of the observation model
- $\boldsymbol{\theta}_{\text{other}}$: other parameters (input uncertainty etc.)

Dynamic HMC sampler (Stan) to obtain posterior samples $\left\{\boldsymbol{\theta}^{(s)}\right\}_{s=1}^{S}$ and $\left\{\mathbf{f}^{(j,s)}\right\}_{s=1}^{S}$

- Sample the full model that includes all covariates / additive kernels

Bayesian inference and covariate relevance

Robust prior specifications

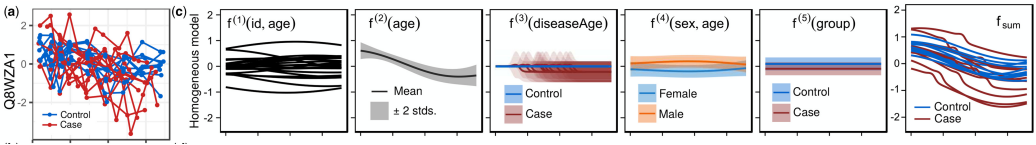
- $\boldsymbol{\theta}_{\text{kernel}}$: all kernel (hyper)parameters
- $\boldsymbol{\theta}_{\text{obs}}$: params. of the observation model
- $\boldsymbol{\theta}_{\text{other}}$: other parameters (input uncertainty etc.)

Dynamic HMC sampler (Stan) to obtain posterior samples $\left\{\boldsymbol{\theta}^{(s)}\right\}_{s=1}^{S}$ and $\left\{\mathbf{f}^{(j,s)}\right\}_{s=1}^{S}$

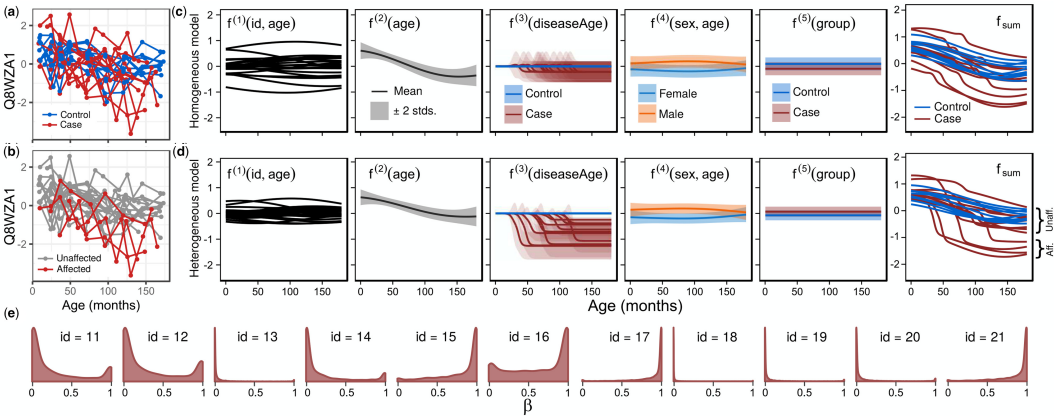
- Sample the full model that includes all covariates / additive kernels

Covariate relevance assessment: decomposition of variance

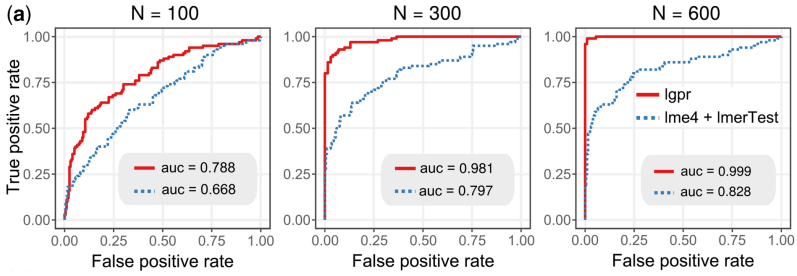
Longitudinal plasma proteomics from a T1D case-control study



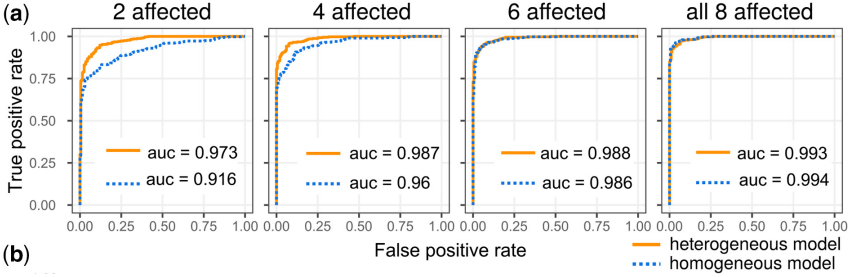
Longitudinal plasma proteomics from a T1D case-control study



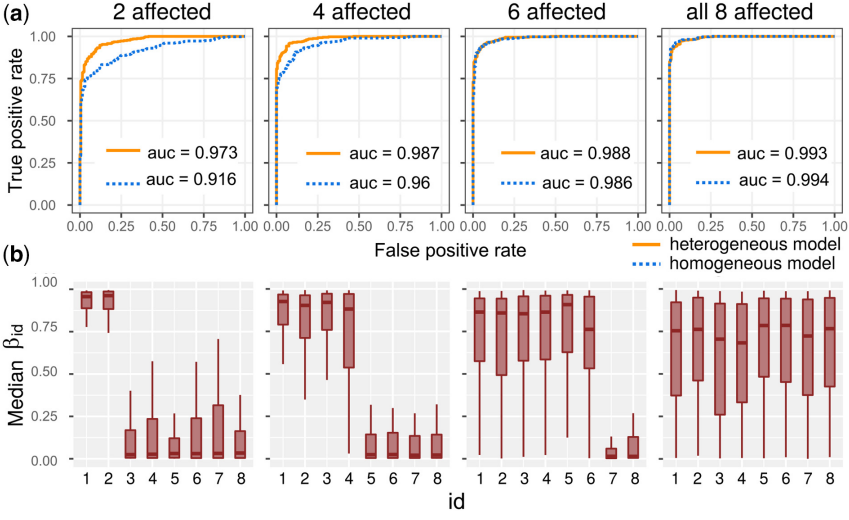
Comparison against LMM



Detection of heterogeneous effects



Detection of heterogeneous effects

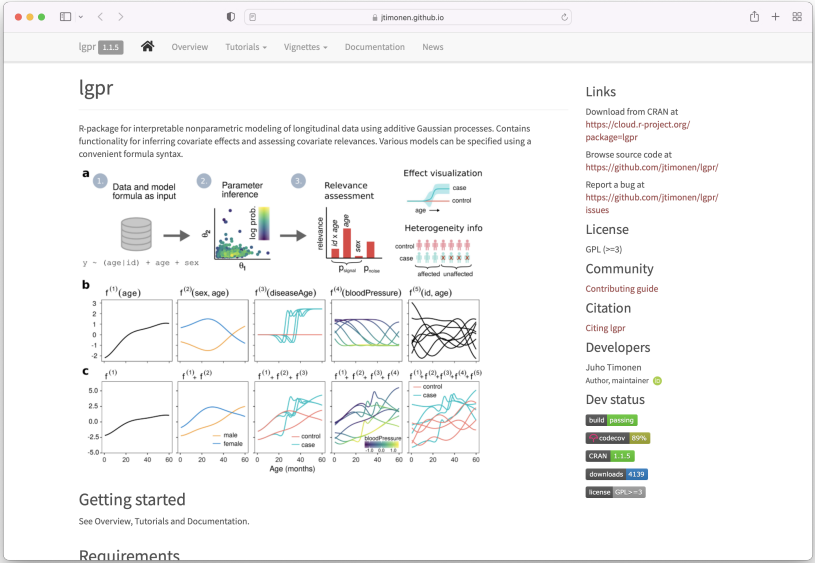


Scalable mixed-domain Gaussian processes

Juho Timonen[1] and Harri Lähdesmäki[1]

[1]Aalto University, Department of Computer Science

Software


jtimonen.github.io
lgpr 1.1.5
Overview
Tutorials
Vignettes
Documentation
News
lgpr
R-package for interpretable nonparametric modeling of longitudinal data using additive Gaussian processes. Contains functionality for inferring covariate effects and assessing covariate relevances. Various models can be specified using a convenient formula syntax.
Data and model formula as input
Parameter inference
Relevance assessment
Effect visualization
Heterogeneity info
y ~ (age|id) + age + sex
Getting started
See Overview, Tutorials and Documentation.
Requirements
Links
Download from CRAN at https://cloud.r-project.org/package=lgpr
Browse source code at https://github.com/jtimonen/lgpr/
Report a bug at https://github.com/jtimonen/lgpr/issues
License
GPL (>=3)
Community
Contributing guide
Citation
Citing lgpr
Developers
Juho Timonen
Author, maintainer
Dev status
build passing
codecov 89%
CRAN 1.1.5
downloads 4139
license GPL (>=3

L-VAE model

Longitudinal Variational Autoencoder

Siddharth Ramchandran[1] **Gleb Tikhonov**[1] **Kalle Kujanpää**[1]
Miika Koskinen[2,3] **Harri Lähdesmäki**[1]

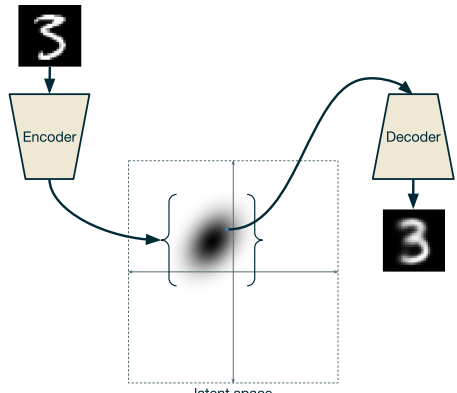
Deep latent variable models

Generative models of the form

$$\mathbf{z} \sim p_\theta(\mathbf{z})$$
$$\mathbf{y} \sim \mathrm{Expfam}(\mathbf{y} \mid d_\psi(\mathbf{z}))$$

Amortized variational inference with $q_\phi(\mathbf{z} \mid \mathbf{y})$ (auto-encoding variational Bayes)

When ϕ and ψ are neural nets, the model is called the variational autoencoder (VAE)



https://ijdykeman.github.io/ml/2016/12/21/cvae.html

Conditional deep latent variable models

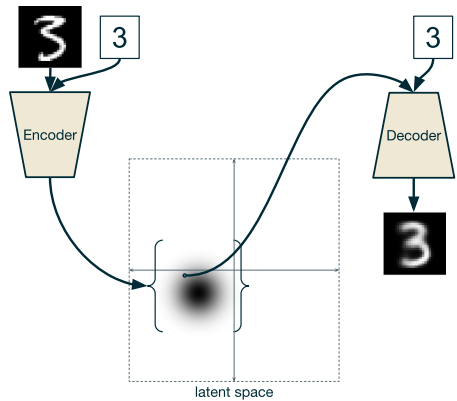
Conditional generative models of the form

$$\mathbf{z} \sim p_\theta(\mathbf{z} \mid \mathbf{x})$$
$$\mathbf{y} \sim \mathrm{Expfam}(\mathbf{y} \mid d_\psi(\mathbf{z}, \mathbf{x}))$$

and amortized variational inference with $q_\phi(\mathbf{z} \mid \mathbf{y}, \mathbf{x})$

Limitations of VAEs and cVAEs

1. The model assumes that the data samples are independent
2. Real-world data contains missing values
3. Real-world data can be multi-modal



https://ijdykeman.github.io/ml/2016/12/21/cvae.html

Deep latent variable models:
Longitudinal, missing and multi-modal data

Multi-output additive GP prior

The L-dimensional latent $\mathbf{z}$ has a vector-valued additive model, conditioned on $\mathbf{x}$

$$\mathbf{z} = \mathbf{f}^{(1)}(\mathbf{x}) + \mathbf{f}^{(2)}(\mathbf{x}) + \ldots + \mathbf{f}^{(R)}(\mathbf{x}) + \mathrm{diag}(\sigma_{z1}^2, \ldots, \sigma_{zL}^2),$$

where each

$$\mathbf{f}^{(r)}(\mathbf{x}) \sim GP(\mathbf{0}, \mathbf{K}^{(r)}(\mathbf{x}, \mathbf{x}'|\theta))$$

with a matrix-valued $(L \times L)$ cross-covariance function $\mathbf{K}^{(r)}(\mathbf{x}, \mathbf{x}'|\theta)$

Multi-output additive GP prior

The L-dimensional latent $\mathbf{z}$ has a vector-valued additive model, conditioned on $\mathbf{x}$

$$\mathbf{z} = \mathbf{f}^{(1)}(\mathbf{x}) + \mathbf{f}^{(2)}(\mathbf{x}) + \ldots + \mathbf{f}^{(R)}(\mathbf{x}) + \mathrm{diag}(\sigma_{z1}^2, \ldots, \sigma_{zL}^2),$$

where each

$$\mathbf{f}^{(r)}(\mathbf{x}) \sim GP(\mathbf{0}, \mathbf{K}^{(r)}(\mathbf{x}, \mathbf{x}'|\theta))$$

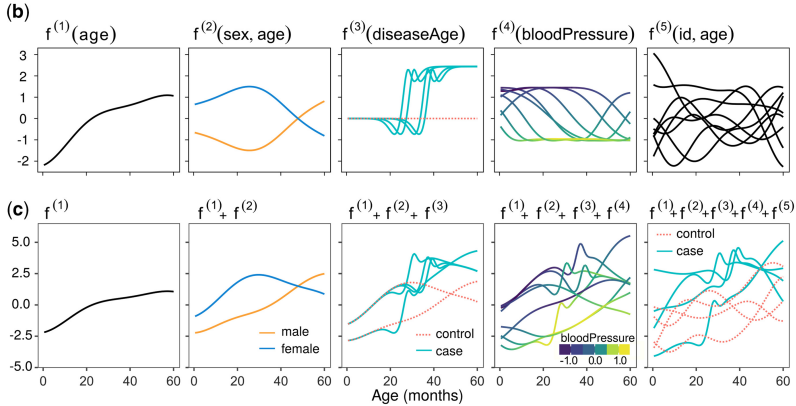
with a matrix-valued $(L \times L)$ cross-covariance function $\mathbf{K}^{(r)}(\mathbf{x}, \mathbf{x}'|\theta)$

Thus

$$\mathbf{z} \sim GP\left(\mathbf{0}, \sum_{r=1}^{R} \mathbf{K}^{(r)}(\mathbf{x}, \mathbf{x}'|\theta) + \mathrm{diag}(\sigma_{z1}^2, \ldots, \sigma_{zL}^2)\right)$$

Kernel functions for longitudinal data

- Shared effects: the exponentiated quadratic (EQ) kernel
- Nonstationary shared effects: EQ kernel with monotonic nonlinear input warping
- Category effects: product of the zero-sum and EQ kernels
- Individual random components



High-dimensional, correlated data

Data:

$$Y = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T = [Y_1^T, \ldots, Y_P^T]^T \qquad D\text{-dimensional data}$$
$$X = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T = [X_1^T, \ldots, X_P^T]^T \qquad Q\text{-dimensional covariates}$$
$$Z = [\mathbf{z}_1, \ldots, \mathbf{z}_N]^T = [\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_L] \qquad L\text{-dimensional latent}$$

The generative model: L-VAE

Conditional probability for the full dataset (X, Y)

$$
\begin{aligned}
p_\omega(Y|X) &= \int_Z p_\psi(Y|Z) p_\theta(Z|X) dZ \\
&= \int_Z \left(\prod_{n=1}^{N} \underbrace{p(\mathbf{y}_n | d_\psi(\mathbf{z}_n))}_{\text{likelihoods}} \right) \underbrace{p_\theta(Z|X)}_{\text{GP prior}} dZ,
\end{aligned}
$$

Auto-encoding variational Bayes for L-VAE

Amortized variational inference:

$$q_\phi(Z|Y) = \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{z}_n|\boldsymbol{\mu}_\phi(\mathbf{y}_n), \text{diag}(\boldsymbol{\sigma}^2_\phi(\mathbf{y}_n))\right)$$

where $\boldsymbol{\mu}_\phi$ and $\boldsymbol{\sigma}^2_\phi$ are neural networks

Auto-encoding variational Bayes for L-VAE

Amortized variational inference:

$$q_\phi(Z|Y) = \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{z}_n|\boldsymbol{\mu}_\phi(\mathbf{y}_n), \text{diag}(\boldsymbol{\sigma}^2_\phi(\mathbf{y}_n))\right)$$

where $\boldsymbol{\mu}_\phi$ and $\boldsymbol{\sigma}^2_\phi$ are neural networks

ELBO objective:

$$\log p_\omega(Y|X) \geq \mathcal{L}(\phi, \psi, \theta; Y, X) \triangleq \underbrace{\mathbb{E}_{q_\phi(Z|Y)}\left[\log p_\psi(Y|Z)\right]}_{\text{easy}} - \underbrace{D_{\text{KL}}(q_\phi(Z|Y)||p_\theta(Z|X))}_{\text{closed form, but slow}}$$

Auto-encoding variational Bayes for L-VAE

Amortized variational inference:

$$q_\phi(Z|Y) = \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{z}_n | \boldsymbol{\mu}_\phi(\mathbf{y}_n), \mathrm{diag}(\boldsymbol{\sigma}^2_\phi(\mathbf{y}_n))\right)$$

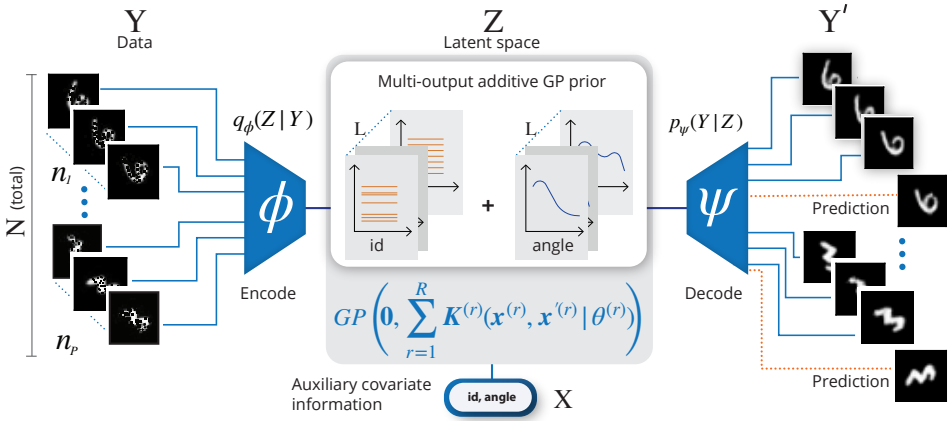
where $\boldsymbol{\mu}_\phi$ and $\boldsymbol{\sigma}^2_\phi$ are neural networks

ELBO objective:

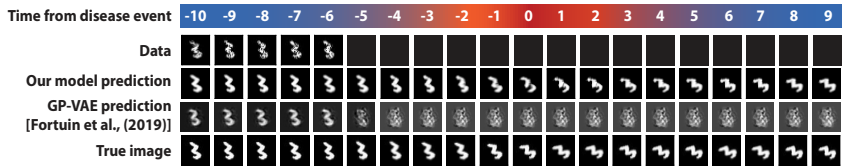
$$\log p_\omega(Y|X) \geq \mathcal{L}(\phi, \psi, \theta; Y, X) \triangleq \underbrace{\mathbb{E}_{q_\phi(Z|Y)}\left[\log p_\psi(Y|Z)\right]}_{\text{easy}} - \underbrace{D_{\mathrm{KL}}(q_\phi(Z|Y)||p_\theta(Z|X))}_{\text{closed form, but slow}}$$

- A novel and provably tighter evidence lower bound for longitudinal GPs ($\rightarrow$ upper bound for the KL)
- A novel mini-batch compatible KL upper bound

Autoencoder view of L-VAE

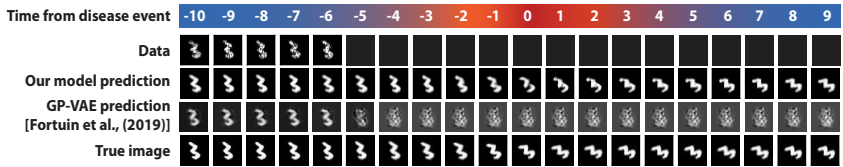


The Health MNIST experiment



- Simulated longitudinal data using MNIST dataset
- Train: $P = 1000$ unique instances ($N = 20000$)
- $Q = 6$ covariates: *id*, *age*, *diseasePresence*, *diseaseAge*, *sex*, and *location*
- Test: 100 additional instances: given time points $[-10, \ldots, -6]$, predict $[-5, \ldots, 9]$

The Health MNIST experiment



- Simulated longitudinal data using MNIST dataset
- Train: $P = 1000$ unique instances ($N = 20000$)
- $Q = 6$ covariates: *id*, *age*, *diseasePresence*, *diseaseAge*, *sex*, and *location*
- Test: 100 additional instances: given time points $[-10, \ldots, -6]$, predict $[-5, \ldots, 9]$

Table 2: MSE from performing future predictions (i.e., from time [-5, 9]) on the Health MNIST dataset. The values are the means and respective standard errors.

Model	Latent dimension	MSE
GPPVAE	64	0.057 ± 0.003
GP-VAE	64	0.059 ± 0.002
VRNN	64	0.049 ± 0.004
BRITS	N/A	0.047 ± 0.004
GRUI-GAN	64	0.053 ± 0.007
L-VAE	8	0.038 ± 0.003
L-VAE	16	0.033 ± 0.0018
L-VAE	**32**	$\mathbf{0.025 \pm 0.0015}$

The Physionet 2012 experiment

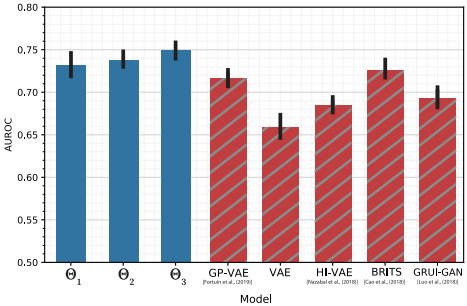
- Predict in-hospital mortality of ICU patients
- Train: $P = 3997$ individuals ($N = 191856$)
- Test: $P = 3993$ individuals
- $D = 35$ features: glucose level, blood pressure, temperature, ...
- $Q = 9$ covariates: id, time, ICUtype, height, weight, age, sex, mortality, mortalityTime

The Physionet 2012 experiment

- Predict in-hospital mortality of ICU patients
- Train: $P = 3997$ individuals ($N = 191856$)
- Test: $P = 3993$ individuals
- $D = 35$ features: glucose level, blood pressure, temperature, ...
- $Q = 9$ covariates: id, time, ICUtype, height, weight, age, sex, mortality, mortalityTime
- Model-based prediction for mortality

$$P_1 = \frac{\exp(L_1)}{\exp(L_0) + \exp(L_1)},$$

where $L_i = \mathcal{L}(\phi, \psi, \theta; Y_*, X_*, \text{mortality} = i)$ for $i = \{0, 1\}$



Randomized clinical trial (RCT) analysis

- Predict appearance of adverse effects (AE) in RCTs (colon cancer treatment)
- $P = 480$ subjects ($N = 6605$)
- $D = 30$ features: lab measurements and vitals
- $Q = 24$ covariates: demographics, adverse effects, medication

Randomized clinical trial (RCT) analysis

- Predict appearance of adverse effects (AE) in RCTs (colon cancer treatment)
- $P = 480$ subjects ($N = 6605$)
- $D = 30$ features: lab measurements and vitals
- $Q = 24$ covariates: demographics, adverse effects, medication
- Model-based prediction for adverse effects

$$P_1 = \frac{\exp(L_1)}{\exp(L_0) + \exp(L_1)},$$

where $L_i = \mathcal{L}(\phi, \psi, \theta; Y_*, X_*, \text{AE} = i)$ for $i = \{0, 1\}$

Adverse effect (AE)	AUC
Skin appendage conditions	0.970
General system disorders nec	0.945
Gastrointestinal signs and symp.	0.900
Gastrointestinal mot. & def.cond.	0.871
White blood cell disorders	0.901
Oral soft tissue conditions	0.966
Neurological disorders nec	0.908
Respiratory disorders nec	0.978
Appetite and GND	0.915
Infections - pathogen unspec.	0.873

Deep latent variable models:
Longitudinal, missing and multi-modal data

Learning conditional VAEs with missing covariates

Many real-world datasets contain missing values

- Missingness in features $\mathbf{y}$
- Missingness in covariates $\mathbf{x}$

Learning conditional VAEs with missing covariates

Many real-world datasets contain missing values

- Missingness in features $\mathbf{y}$
- Missingness in covariates $\mathbf{x}$

Goal: learn conditional VAE models from partially observed datasets that contain missing values also in the auxiliary covariates $\mathbf{x}$

For each data sample $(\mathbf{x}, \mathbf{y})$ any subset of the covariates $\mathbf{x}$ and observed variables $\mathbf{y}$ may be missing completely at random (MCAR)

$$\mathbf{x} = (\mathbf{x}^o, \mathbf{x}^u) \qquad \mathbf{y} = (\mathbf{y}^o, \mathbf{y}^u)$$
$$X = (X^o, X^u) \qquad Y = (Y^o, Y^u)$$

The model

Augment the generative model with a prior distribution for $\mathbf{x}$, $p_\lambda(\mathbf{x})$

$$p_\omega(\mathbf{y}, \mathbf{z}, \mathbf{x}) = p_\psi(\mathbf{y} \mid \mathbf{z}) p_\theta(\mathbf{z} \mid \mathbf{x}) p_\lambda(\mathbf{x})$$

Covariates $\mathbf{x}$ contain both discrete and continuous variables

- Amortised variational approximation for $\mathbf{x}^u$
- Maximize the ELBO while simultaneously marginalising uncertainty associated with the missing covariates

The model

Augment the generative model with a prior distribution for $\mathbf{x}$, $p_\lambda(\mathbf{x})$

$$p_\omega(\mathbf{y}, \mathbf{z}, \mathbf{x}) = p_\psi(\mathbf{y} \mid \mathbf{z}) p_\theta(\mathbf{z} \mid \mathbf{x}) p_\lambda(\mathbf{x})$$

Covariates $\mathbf{x}$ contain both discrete and continuous variables

- Amortised variational approximation for $\mathbf{x}^u$
- Maximize the ELBO while simultaneously marginalising uncertainty associated with the missing covariates

The ELBO with missing covariate

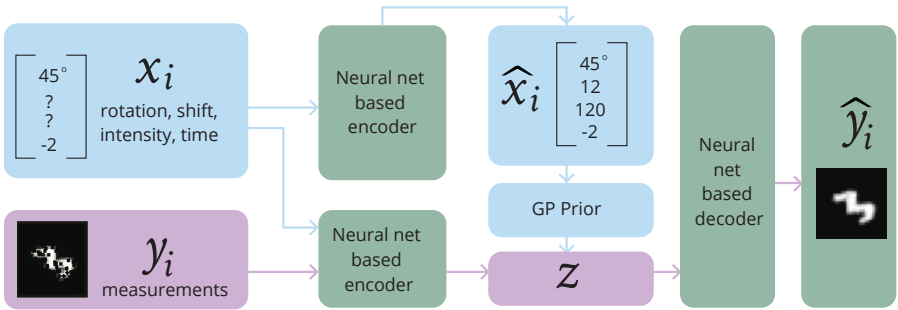
$$\begin{aligned}
&\log p_\omega(Y^{\mathrm{o}} | X^{\mathrm{o}}) \geq \mathcal{L}(\phi, \psi, \theta, \lambda; Y^{\mathrm{o}}, X^{\mathrm{o}}) \\
&\triangleq \mathbb{E}_q[\log p_\psi(Y^{\mathrm{o}} | Z)] \\
&\quad - \underbrace{\mathrm{KL}[q_\phi(Z, X^{\mathrm{u}} | Y^{\mathrm{o}}, X^{\mathrm{o}}) || p_{\theta,\lambda}(Z, X^{\mathrm{u}} | X^{\mathrm{o}})]}_{D_{\mathrm{KL}}}
\end{aligned}$$

where

$$\begin{aligned}
D_{\mathrm{KL}} = &\mathbb{E}_q \left[\mathrm{KL}[q_\phi(Z | Y^{\mathrm{o}}, X^{\mathrm{o}}) || p_\theta(Z | X^{\mathrm{u}}, X^{\mathrm{o}})] \right] \\
&+ \mathrm{KL}[q_\phi(X^{\mathrm{u}} | X^{\mathrm{o}}) || p_\lambda(X^{\mathrm{u}})]
\end{aligned}$$

Mini-batch compatible scalable computation

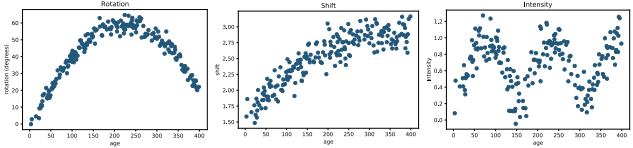
VAE model overview



Time-series MNIST dataset

Manipulated digits from the MNIST dataset

- digit rotation
- shift along diagonal
- image intensity
- [time]



NLL for test predictions

Method	Dataset	Missing %				
		5%	10%	20%	30%	40%
Temporal L-VAE with mean impute	Dataset 3	0.19 ± 0.008	0.27 ± 0.008	0.39 ± 0.02	0.42 ± 0.01	0.56 ± 0.03
Temporal L-VAE with KNN impute	Dataset 3	0.17 ± 0.005	0.25 ± 0.009	0.35 ± 0.01	0.47 ± 0.02	0.58 ± 0.03
Temporal L-VAE with HI-VAE impute	Dataset 3	0.14 ± 0.004	0.21 ± 0.007	0.29 ± 0.009	0.42 ± 0.01	0.51 ± 0.03
Temporal L-VAE with our method †	Dataset 3	$\mathbf{0.12 \pm 0.003}$	$\mathbf{0.18 \pm 0.002}$	$\mathbf{0.23 \pm 0.004}$	$\mathbf{0.35 \pm 0.02}$	$\mathbf{0.46 \pm 0.03}$
Temporal L-VAE with our method	Dataset 3	$\mathbf{0.12 \pm 0.002}$	$\mathbf{0.19 \pm 0.003}$	$\mathbf{0.25 \pm 0.006}$	$\mathbf{0.38 \pm 0.01}$	$\mathbf{0.44 \pm 0.02}$
Temporal L-VAE with oracle	Dataset 3	0.11 ± 0.001	0.16 ± 0.003	0.21 ± 0.007	0.28 ± 0.008	0.37 ± 0.01

MSE in test set covariate imputation

Method	Dataset	Missing %				
		5%	**10%**	**20%**	**30%**	**40%**
Mean impute	Dataset 3	0.75	0.77	0.95	0.97	1.25
KNN impute	Dataset 3	0.71	0.74	1.025	1.012	1.35
Temporal L-VAE with our method	Dataset 3	$\mathbf{0.21 \pm 0.02}$	$\mathbf{0.22 \pm 0.04}$	$\mathbf{0.29 \pm 0.06}$	$\mathbf{0.35 \pm 0.04}$	$\mathbf{0.68 \pm 0.06}$

Randomized clinical trial (RCT) analysis

Longitudinal data from a RCT (prostate cancer treatment) observed over a period of $\sim$2 years

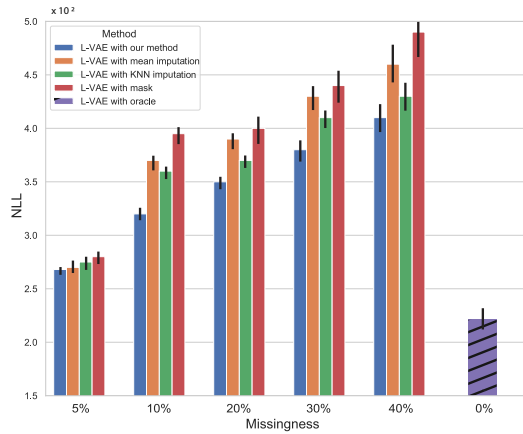
- 184 patients, 1287 data samples
- $D = 28$ lab measurements as well as vital signs (Y)
- $Q = 23$ patient-specific auxiliary covariates (X)

Randomized clinical trial (RCT) analysis

Longitudinal data from a RCT (prostate cancer treatment) observed over a period of $\sim$2 years

- 184 patients, 1287 data samples
- $D = 28$ lab measurements as well as vital signs (Y)
- $Q = 23$ patient-specific auxiliary covariates (X)

Evaluate NLL for test set predictions



A Parkinson's Progression Markers Initiative Dataset (Marek et al., 2011)

Longitudinal data from a observational studies

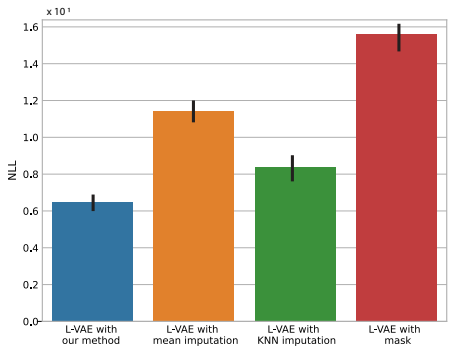
- Approx. 5 year follow-up study
- 545 patients: 371 PD, 174 healthy
- Total 3135 measurements
- $D = 42$ features in Y: cognitive tests, DaTSCAN, cerebrospinal fluid results, bio-specimen, etc. with $\sim 4.8\%$ missing values
- $Q = 7$ auxiliary covariates (X) with missingness between 2% and 32%

A Parkinson's Progression Markers Initiative Dataset (Marek et al., 2011)

Longitudinal data from a observational studies

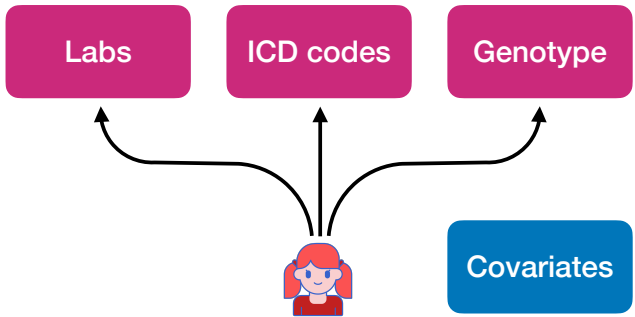
- Approx. 5 year follow-up study
- 545 patients: 371 PD, 174 healthy
- Total 3135 measurements
- $D = 42$ features in Y: cognitive tests, DaTSCAN, cerebrospinal fluid results, bio-specimen, etc. with $\sim 4.8\%$ missing values
- $Q = 7$ auxiliary covariates (X) with missingness between 2% and 32%

Evaluate NLL for test set predictions



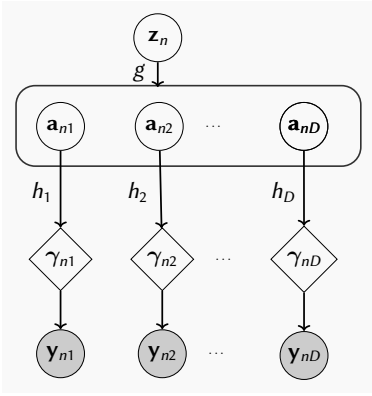
Deep latent variable models:
Longitudinal, missing and multi-modal data

Biomedical data sets are often multi-modal



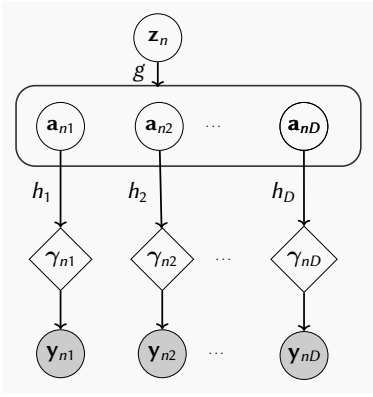
L-VAE model for multi-modal / heterogeneous data (HL-VAE)

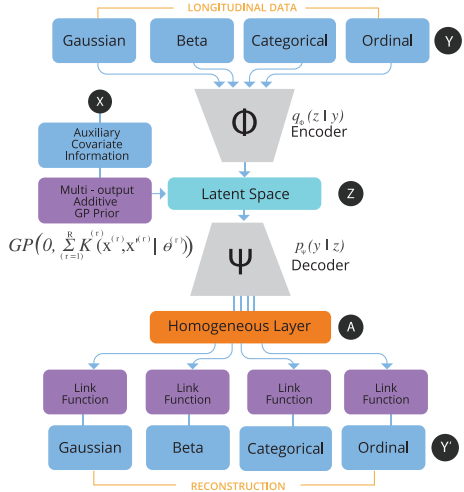
We utilize ideas from (Nazabel et al., 2020) for modeling heterogeneous data



L-VAE model for multi-modal / heterogeneous data (HL-VAE)

We utilize ideas from (Nazabel et al., 2020) for modeling heterogeneous data





A Parkinson's Progression Markers Initiative Dataset (Marek et al., 2011)

- Approx. 5 year follow-up study
- 545 patients: 371 PD, 174 healthy
- Total 3135 measurements
- $D = 80$ dimensional measurements (Y)
- $Q = 7$ auxiliary covariates (X)

	Configuration
Gaussian Distribution	8
LogNormal Distribution	12
Poisson Distr.	12
Ordinal	12
Categorical	36

A Parkinson's Progression Markers Initiative Dataset (Marek et al., 2011)

- Approx. 5 year follow-up study
- 545 patients: 371 PD, 174 healthy
- Total 3135 measurements
- $D = 80$ dimensional measurements (Y)
- $Q = 7$ auxiliary covariates (X)

	Configuration
Gaussian Distribution	8
LogNormal Distribution	12
Poisson Distr.	12
Ordinal	12
Categorical	36

Non Categorical

Dataset	L-VAE ↓	HL-VAE ↓
Future Prediction		
10% Missing	0.099 ± 0.003	$\mathbf{0.086 \pm 0.002}$
20% Missing	0.097 ± 0.001	$\mathbf{0.086 \pm 0.003}$
30% Missing	0.1 ± 0.001	$\mathbf{0.091 \pm 0.002}$
40% Missing	0.102 ± 0.001	$\mathbf{0.088 \pm 0.004}$
50% Missing	0.105 ± 0.002	$\mathbf{0.094 \pm 0.002}$
Test Prediction		
10% Missing	0.093 ± 0.004	$\mathbf{0.079 \pm 0.003}$
20% Missing	0.091 ± 0.002	$\mathbf{0.080 \pm 0.003}$
30% Missing	0.094 ± 0.002	$\mathbf{0.087 \pm 0.003}$
40% Missing	0.095 ± 0.001	$\mathbf{0.085 \pm 0.001}$
50% Missing	0.099 ± 0.001	$\mathbf{0.091 \pm 0.003}$

(b) NRMSE

Acknowledgements

Aalto Univ.

- **Lu Cheng**
- Kalle Kujanpää
- Otto Lönnroth
- Henrik Mannerström
- **Mine Ögretir**
- Dimitris Papatheodorou
- **Siddharth Ramchandran**
- **Gleb Tikhonov**
- **Juho Timonen**
- Tommi Vatanen
- **Aki Vehtari**

Univ. of Helsinki

- Miika Koskinen

Univ. of Turku

- Niina Lietzen
- Riitta Lahesmaa

Bayer Pharmaceuticals

- Pekka Tiikkainen
- Jussi Leinonen

Funding

- Academy of Finland
- Bayer Pharmaceuticals

References

- Cheng L, Ramchandran S, Vatanen T, Lietzen N, Lahesmaa R, Vehtari A and Lähdesmäki H, An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data, *Nature Communications*, Vol. 10, No. 1798, 2019.
- Timonen J, MannerstrÃ¶m H, Vehtari A, Lähdesmäki H, lgpr: An interpretable nonparametric method for inferring covariate effects from longitudinal data, *Bioinformatics*, Vol. 37, No. 13, pp. 1860-1867, 2021.
- Ramchandran S, Tikhonov G, Kujanpää K, Koskinen M, Lähdesmäki H, Longitudinal variational autoencoder, *AISTATS*, 2021.
- Ramchandran S, Tikhonov G, Lönnroth O, Tiikkainen P, Lähdesmäki H Learning conditional variational autoencoders with missing covariates, `https://arxiv.org/abs/2203.01218`
- Ögretir M, Ramchandran S, Papatheodorou D, Lähdesmäki H A variational autoencoder for heterogeneous longitudinal data, `https://arxiv.org/abs/2204.09369`