Acceleration of quantitative modeling of fMRI data

DSI Capstone project Spring 2023

04/24/2023

PRESENTED BY
Yuting Mei

ADVISED BY
Dr. Hakmook Kang

Summary

1

Introduction

2

Methodology

3

Result

4

Limitations & Future work

Introduction

Background & Challenging & Dataset & Goal

Background

•High variation in resting-state functional connectivity(rs-FC) and structural connectivity (SC) measures due to limited sample size
•MCMC:
•pros: a flexible and comprehensive way to model complex probability distributions
•Cons: long computational time in simulating high dimensional data, saving computational cost to increase speed(over 15 days for one subject modeling)

Challenging

•the number of possible combinations of parameters grows exponentially in high dimensional case, making it computationally intractable to explore the entire parameter space
•Subsampling for high dimensional spatial-temporal dataset is not suitable for MCMC, hard for parallelizing dataset

BACK TO AGENDA

1

Dataset

Dataset for testing computational time: depressed participants were enrolled at Duke University Medical Center

rs-fMRI data from Diffusion Weighted MRI (DW-MRI), subject id 8007: temporal dataset with shape of 26679 by 150

Spatial distance matrix with shape of 300 by 300 between pairs of ROIs90 ROIs(spatial dataset)

•Functional and structural connectivity dataset with shape of 45 by 4006

Goal

accelerating computational time of estimation of rs-FC for enabling including more samples for reduce variance of result

Methodology

Bayesian modeling procedure & GPLVM &
Cholesky decomposition & Jax & MCMC

Methodology
- Bayesian modeling procedure

Modeling: Bayesian Double Fusion Model(Bayesian hierarchical spatio- temporal model that incorporates structural connectivity (SC) into estimating FC in voxels of region of interests(ROI)), Specify prior distribution over parameters and likelihood function

$$\mathrm{Y}_c(t) = \boldsymbol{\beta}_c + \mathbf{b}_c + \mathbf{d}_c + \epsilon_c(t)$$

$$\begin{aligned} \beta_c &\sim N\left(0,\ \sigma^2_{\beta_c}\right) \\ \mathbf{b}_c &\sim N(0,\ \Sigma_{b_c}) \\ \mathrm{d} &\sim N(0,\ \Sigma_{\mathrm{d}}) \\ \epsilon_{cv}(t) &\sim N\left(0,\ \sigma^2_{\epsilon_{cv}}/\left(1-\phi^2_{cv}\right)\right) \end{aligned}$$

$\mathbf{b}_c$ distance- dependent spatial correlation between voxels inside the same ROI

Σ_{d} Linear combination of SC and FC

d: correlation matrix of FC among ROIs

ϵ_{cv} Temporal correlation AR(1)

Compute posterior distribution over parameters: approximating the posterior distribution of model parameters using MCMC

Make predictions: the posterior distribution can be used to estimate the probability or distribution of the outcome variable given the new data.

Optimize hyperparameters: further tuning parameter to improve model fitting

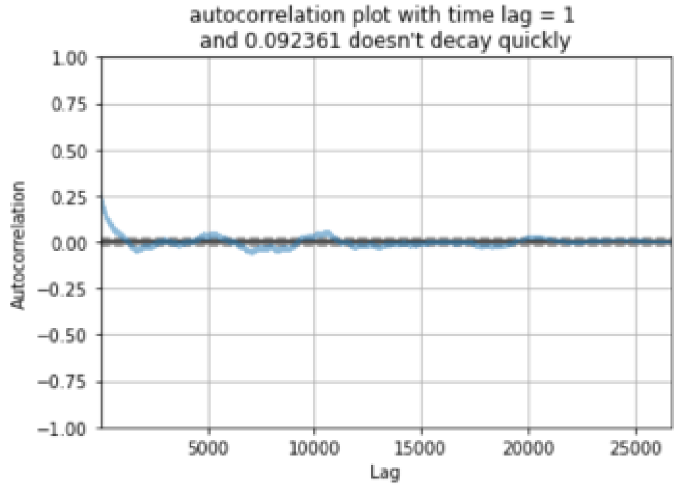
Methodology - GPLVM

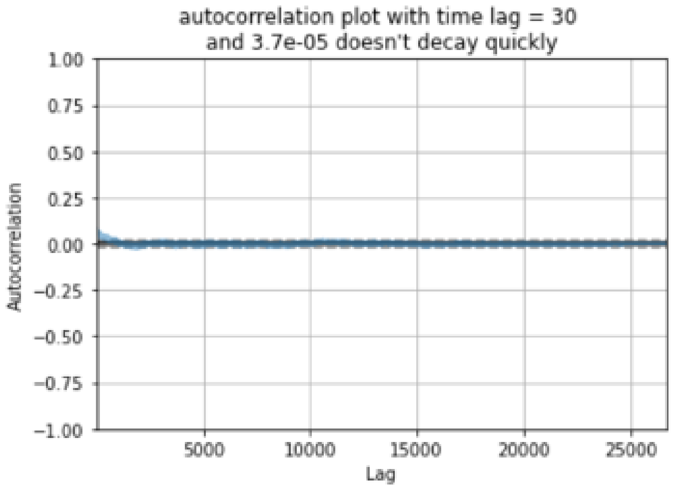
Definition

Learn low dim from high dim:

assumes the original observed data is generated from a low dimensional latent space through a nonlinear mapping function that is corrupted by Gaussian noise

$$\mathbf{X}^{N\times Q} \longleftarrow \mathbf{Y}^{N\times D}, \quad Q \ll D$$





General procedure of GPLVM

- **Define the prior distributions for the kernel hyperparameters**

- Inverse Gamma distribution for lengthscale in temporal dataset(shape = 3, sacle = 0.5), gamma distribution for lengthscale in spatial dataset(moderate variance ~ 10 in each column, choose relatively small alpha and bigger beta for more variation)

- **Define kernel function for covariance matrix(N by N) for latent variable**

- Exponential kernel for temporal dataset(temporal datasets with stationary properties), Matern32 for spatial dataset(nice property with positive definite and differentiable)

- **Define the likelihood, specify latent space(temporal dataset: 5 to 125, spatial dataset: 5 to 300)**
- **Sampling posterior distribution using MCMC with BFGS as optimization method**

- Subsampling for temporal dataset(subsample 1000, really big for 26679 times 26679!)

- **AIC, BIC as model performance evaluation, use for dimension reduction for acceleration if good else keep the visualization**

BACK TO AGENDA

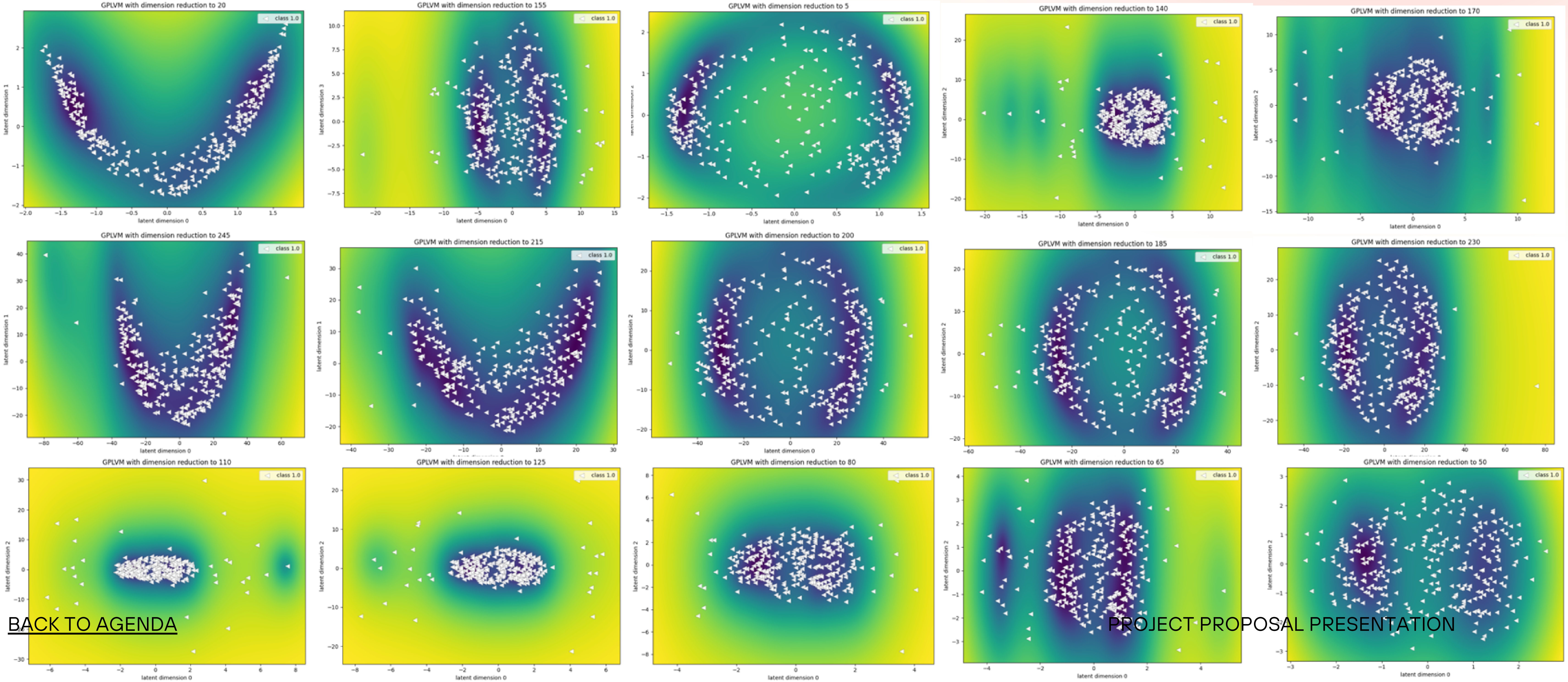
GPLVM on Spatial dataset

Model might not able to fully capture the underlying structure or variation, need more complex one

Banana shape:
Nonlinear relationship(curvature)
Linear relationship(elongated/stretched out direction)

Donut shape(circular or toroidal structure):
periodic or cyclic patterns

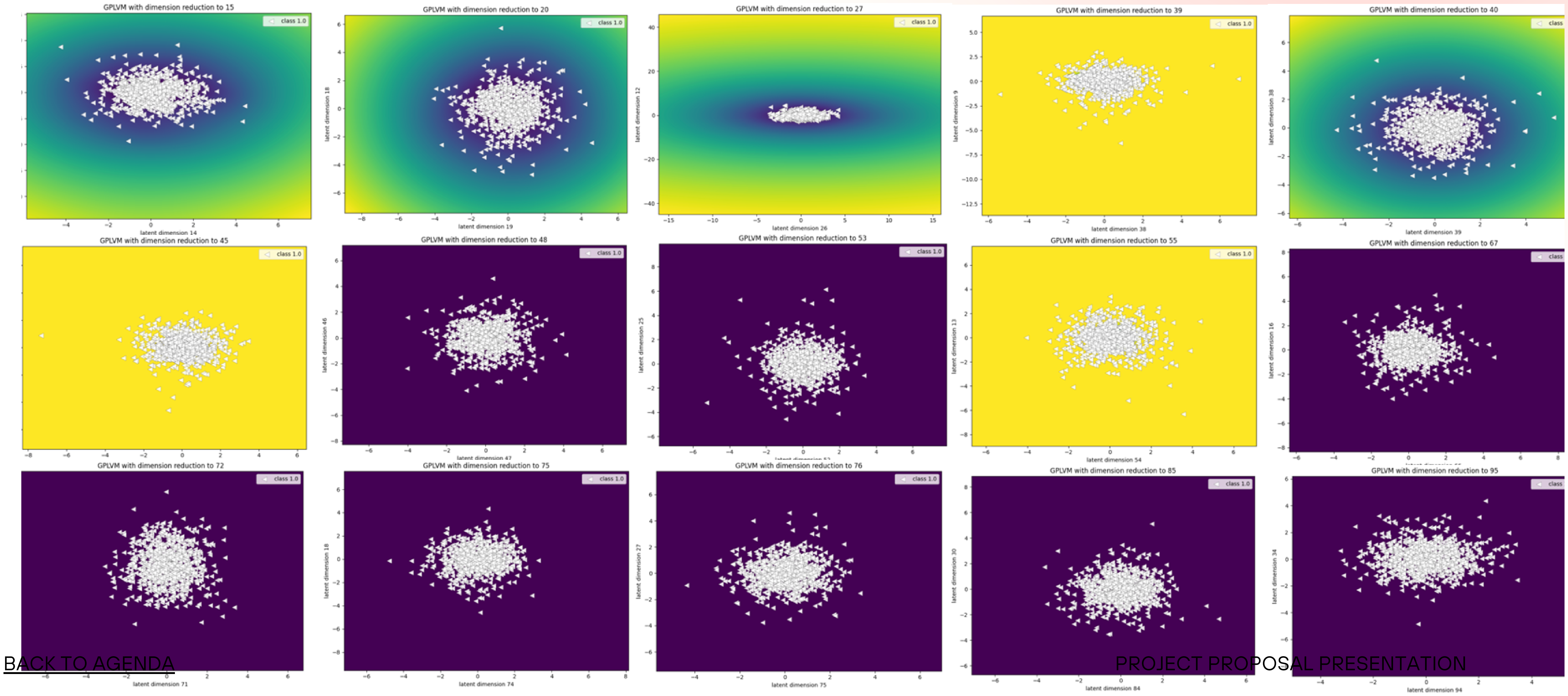
Clusters with dark contours in centroid



GPLVM on Temporal dataset

Model does not fit well

Data densely gather in one center, mostly either all dark or light contours:
the model is not able to distinguish between different classes or patterns



Methodology - Cholesky decomposition, Jax

- **Cholesky decomposition**
- decompose a positive-definite matrix(covariance matrix in prior distribution of parameters) into product of a lower triangular matrix and its transpose
- transform the multivariate normal distribution into a set of independent standard normal variables
- **Jax**
- high-performance library for numerical computing that specializes in automatic differentiation and acceleration of Pymc5
- Take advantage of GPU

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}}_{R} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \underbrace{L_u L_u^T}_{R} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

Methodology - MCMC

Metropolis-Hastings:
inefficient at exploring the complex, nonlinear parameter space caused by random walk

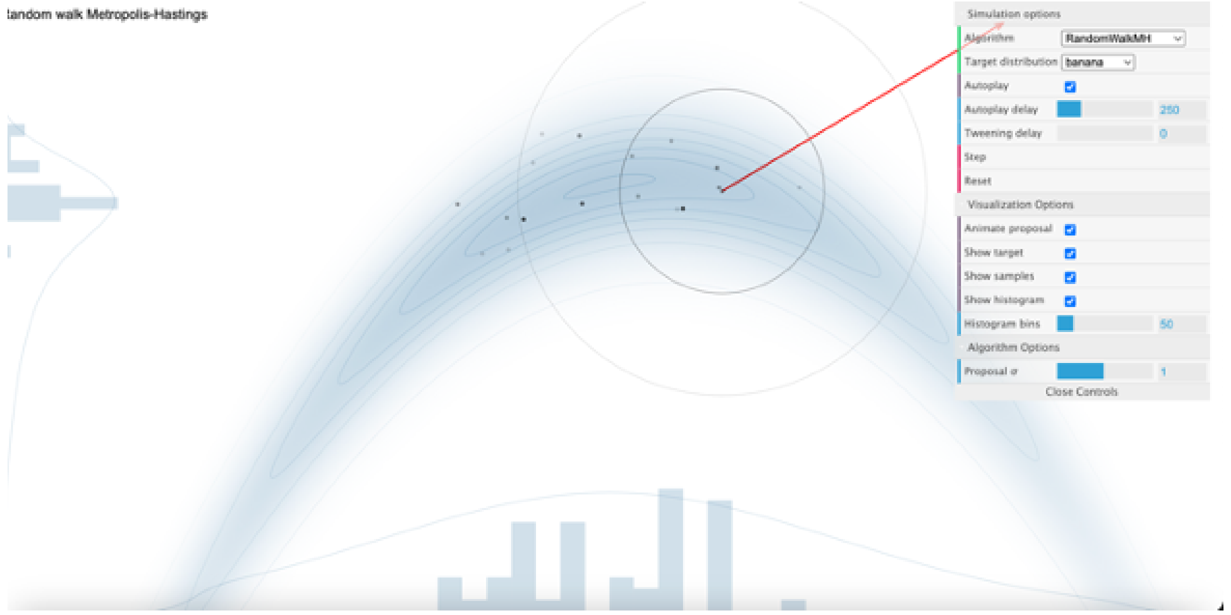
Target distribution of banana shape

Hamiltonian MC:

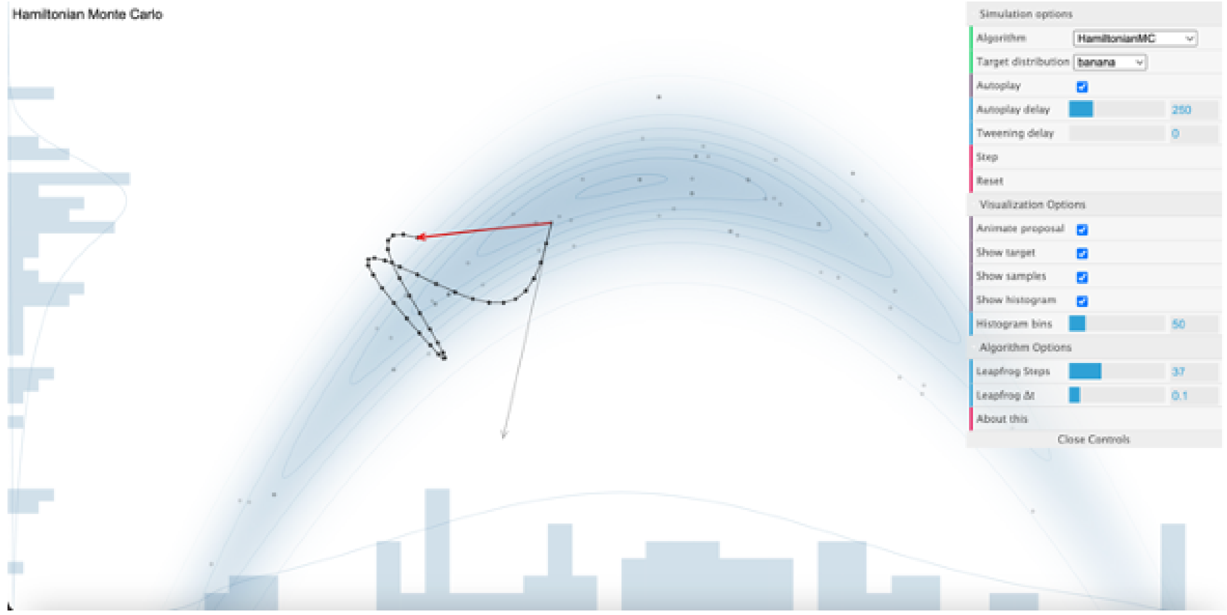
- gradient-based approach, move efficiently with avoiding random walk
- Ability to handle high-dimensional models
- U in trajectory where the acceptance probability is zero

No-U-Turn Sampler:

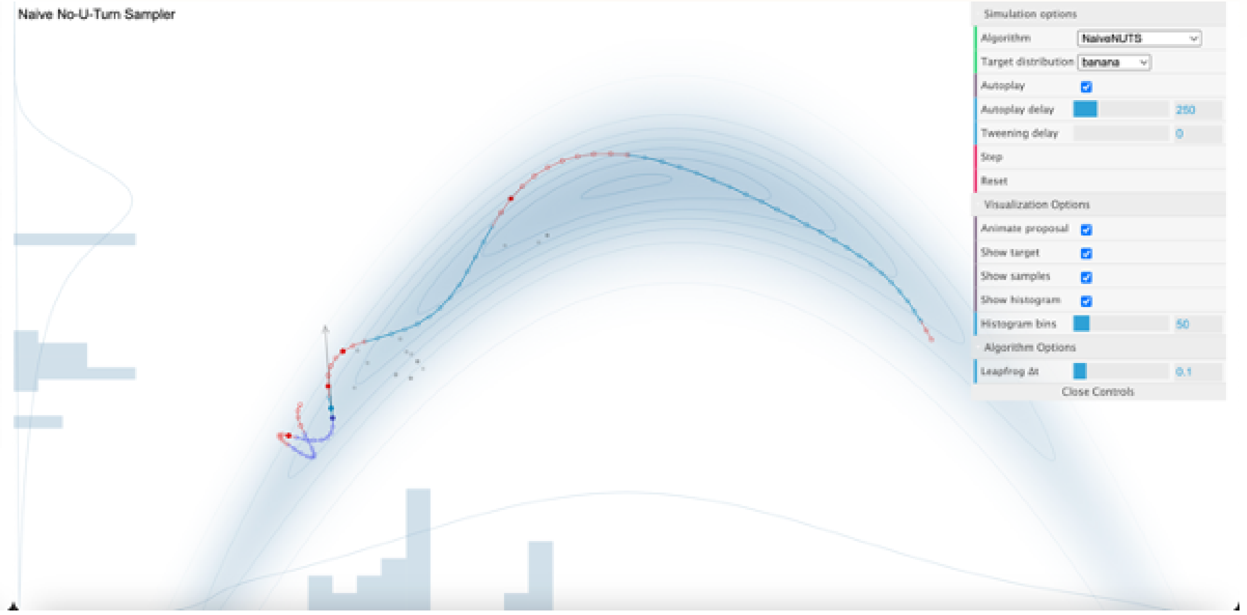
- efficiently explore complex posterior distributions with irregular shapes and high correlations between variables, faster convergence
- Reduce additional hyperparameter setting compared with HMC
- Doubling step or backtrack



Classic MH



Hamiltonian MC



No-U-Turn Sampler

https://chi-feng.github.io/mcmc-demo/app.html#RandomWalkMH

Result

Result

Accelerating computation time from over 200 hours(about 15 days actually) to 1.5 hours on same dataset using PyMC5 with both hardware and built in Gaussian Processes and Cholesky decomposition for sparsity generation

whole-brain saFC (∼100 regions of interest (ROIs))	PyMC3 + CPU	PyMC5 + Jax + GPU	PyMC5 + 5 GPU + 256 CPU	PyMC5 + GP + Jax + GPU	PyMC5 + GP + 256CPU	PyMC5 + GP + Jax + (8GPUs + 256 CPUs)
time estimation for computation	$\sim 200:00:00$	$\sim 144:00:00$	$\sim 16:57:07 + \sim 24:00:00$ compiling	$\sim 18:00:00$	$\sim 5:48:14$	$1:32:19$

BACK TO AGENDA

Limitations & Future work

Limitations & Future work

- Cholesky decomposition: $O(N^3)$ time and $O(N^2)$ memory

- Try low-rank approximation algorithm of the covariance matrix(only need vector operations instead of matrix)

- Clustering for spatial dataset before modeling by spatial and temporal proximity to reduce the number of data points that need to be modeled in MCMC

- Future application to estimation of rs-FC in data from Children with Autism Spectrum Disorder, monitoring convergence and analyzing correlation of saFC values by lasso regression

Thanks for listening!