Stony Brook University

Class-imbalanced Classifiers using Ensembles of Gaussian Processes and Gaussian Process Latent Variable Models

Liu Yang*, Cassandra Heiselman†, J. Gerald Quirk†, Petar M. Djurić*

* Department of Electrical & Computer Engineering, Stony Brook University,
† Department of Obstetrics, Gynecology and Reproductive Medicine, Stony Brook University Hospital, Stony Brook, NY, 11794, USA

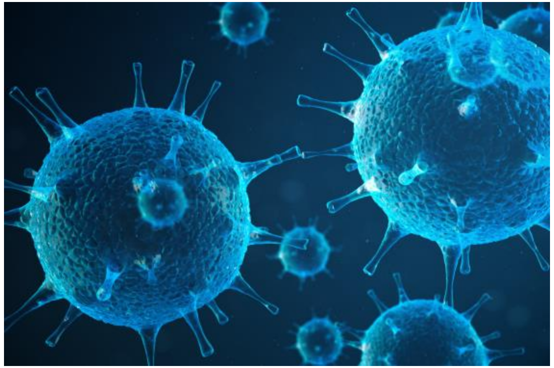
IEEE ICASSP 2021

FAR BEYOND

Overview

- ❖Motivation
- ❖Problem Formulation
- ❖Gaussian Process Latent Variable Model (GPLVM) for Ensemble Classification
- ❖Experiments
- ❖Conclusions

Motivation

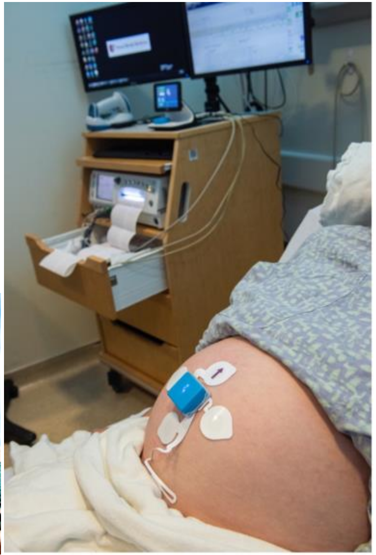


Picture sources:
https://www.helcim.com/article/overview-credit-card-transaction-types/
https://nano-magazine.com/news/2020/8/25/nanoengineered-biosensors-for-early-disease-detection

Motivation

In our project, “Machine Learning Methods for Revealing the Wellbeing of Fetuses”, we face the severe imbalanced fetal heart rate (FHR) recordings.

Only 0.1% FHR tracings are classified into abnormal group.



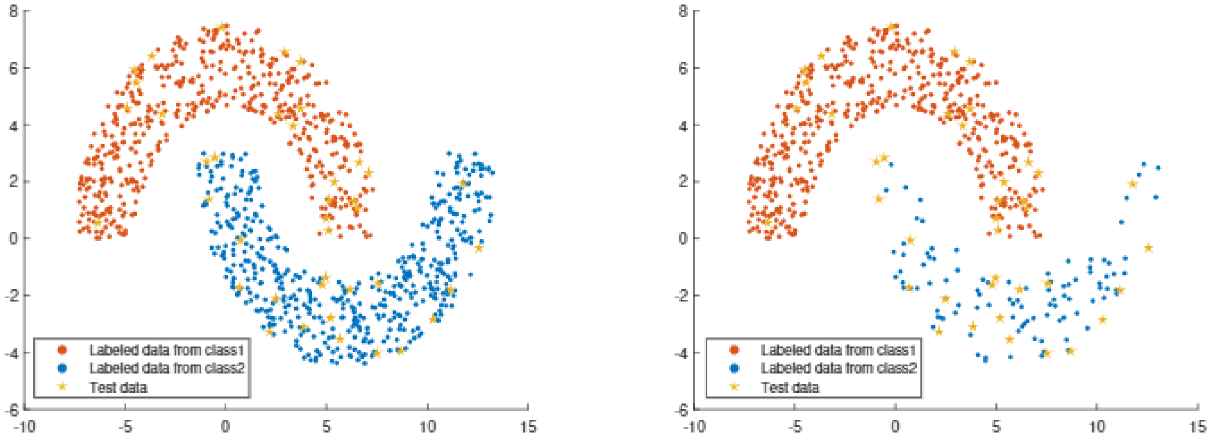
Picture source: https://www.stonybrook.edu/commcms/electrical/research/2021/djuric.php

Problem Formulation

The training set contains a majority class and a minority class:

$$\mathcal{D} = \mathcal{C}_1 \cup \mathcal{C}_2 \qquad \mathcal{C}_1 = \{(\mathbf{x}_i, +1) | i = 1, 2, \ldots, n_{c_1}\}$$

$$\mathcal{C}_2 = \{(\mathbf{x}_j, -1) | j = 1, 2, \ldots, n_{c_2}\} \qquad n_{c_1} \gg n_{c_2}$$



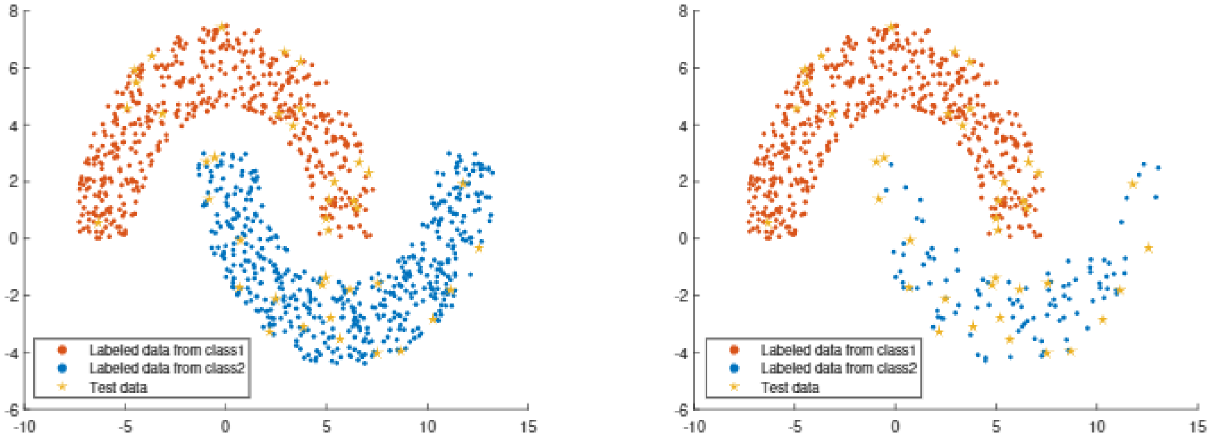
(a) (b)

Problem Formulation

The training set contains a majority class and a minority class:

Rewritten: $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ $\mathbf{X} \in \mathbb{R}^{dx \times n}$ $\mathbf{y} \in \mathbb{R}^{n}$ $n = n_{c_1} + n_{c_2}$

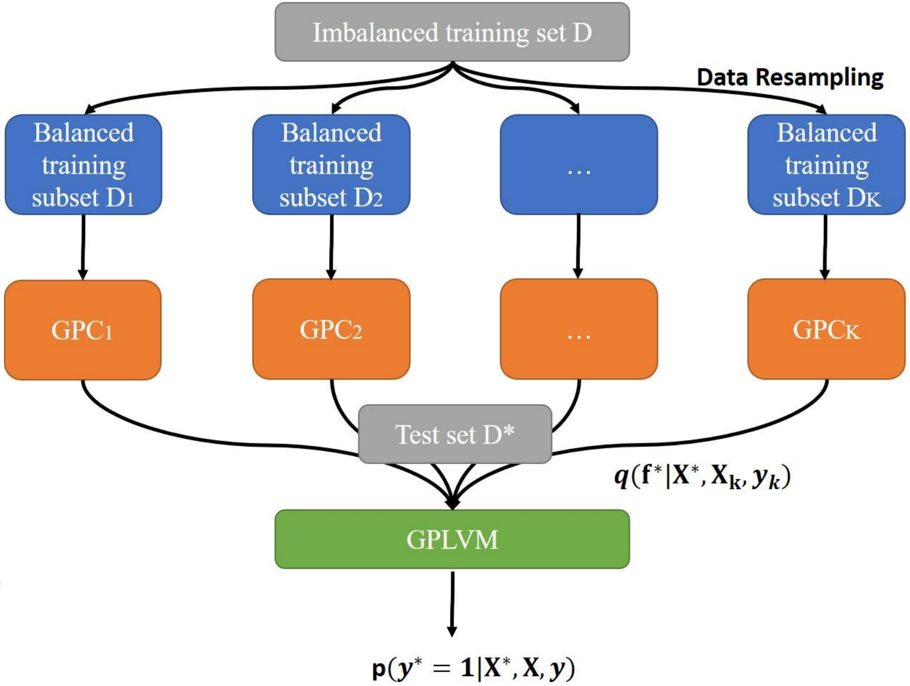
Similarly, the test set: $\mathcal{D}^* = \{\mathbf{X}^*, \mathbf{y}^*\}$ $\mathbf{X}^* \in \mathbb{R}^{dx \times n^*}$ $\mathbf{y}^* \in \mathbb{R}^{n^*}$



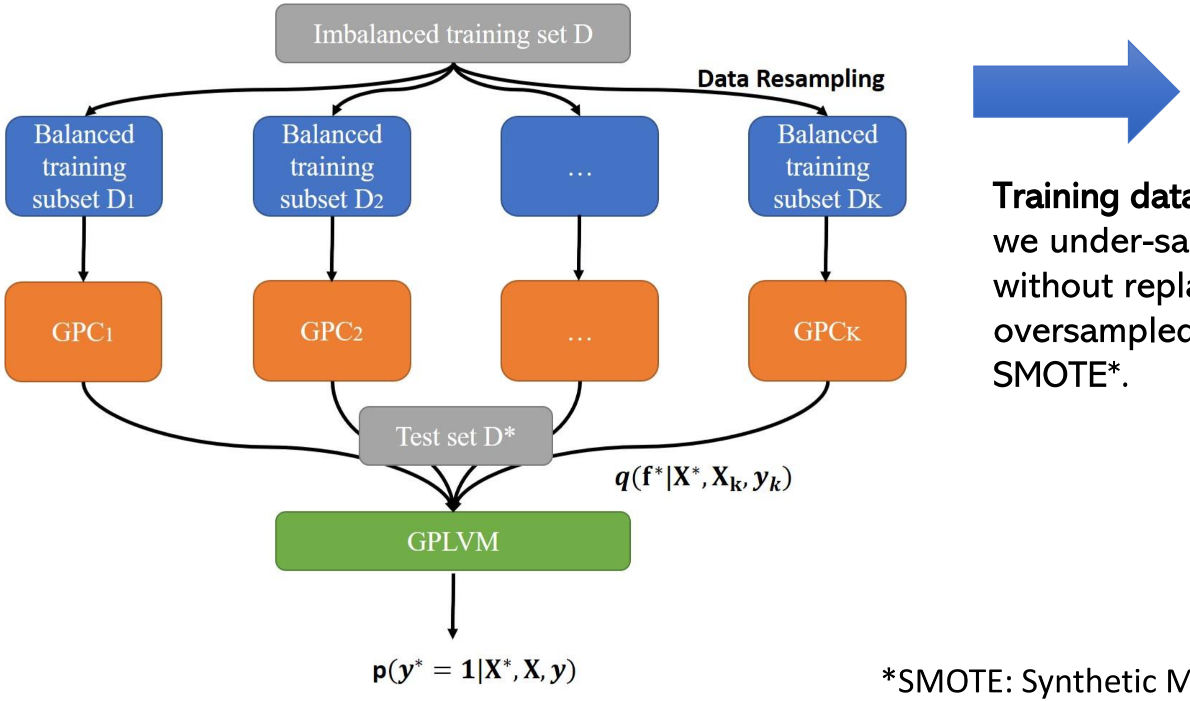
(a) (b)

GPLVM for Ensemble Classification

One of the popular strategies to reduce the effect of performance distortion towards the majority class in the training process is ensemble clearing.



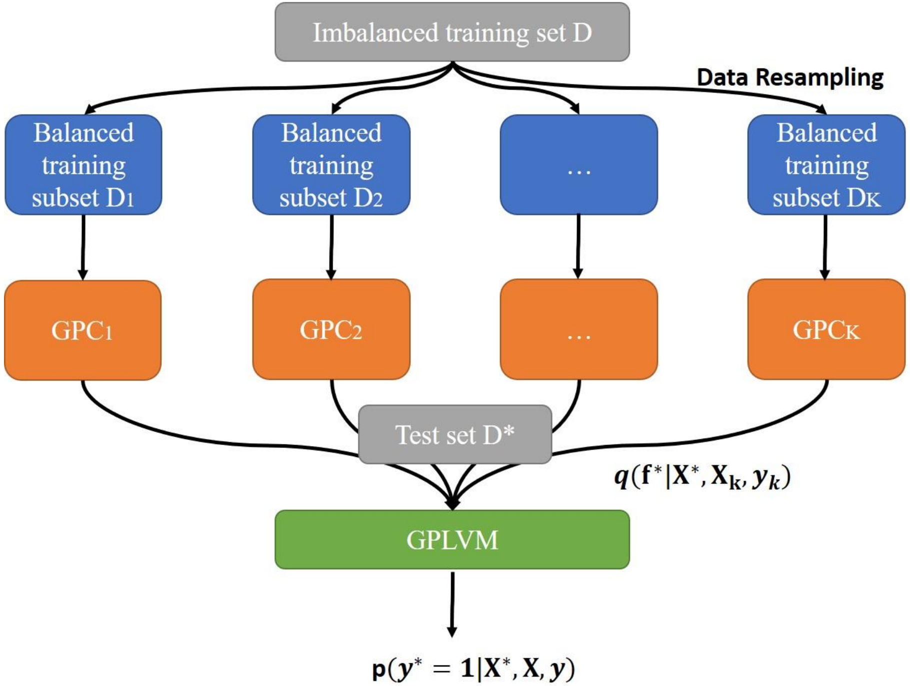
GPLVM for Ensemble Classification



Training data resampling
we under-sampled the majority class without replacement and oversampled the minority class by applying SMOTE*.

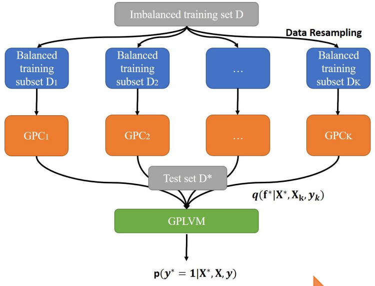
*SMOTE: Synthetic Minority Over-sampling Technique

GPLVM for Ensemble Classification



Stony Brook University

GPLVM for Ensemble Classification

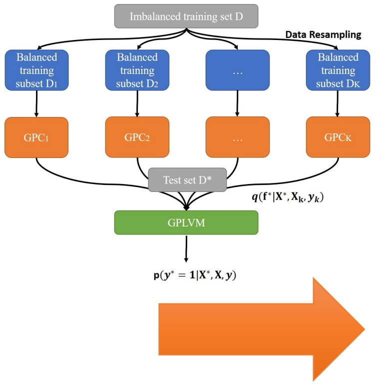


For each branch

The posterior

$$p(\mathbf{f}|\mathbf{X}_k, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{f})p(\mathbf{f}|\mathbf{X}_k)}{p(\mathbf{y}_k|\mathbf{X}_k)}$$

GPLVM for Ensemble Classification



For each branch

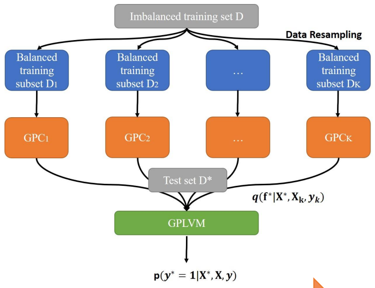
The posterior

$$p(\mathbf{f}|\mathbf{X}_k, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{f})p(\mathbf{f}|\mathbf{X}_k)}{p(\mathbf{y}_k|\mathbf{X}_k)}$$

The predictive distribution

$$p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \int p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{f})p(\mathbf{f}|\mathbf{X}_k, \mathbf{y}_k)d\mathbf{f}$$

GPLVM for Ensemble Classification



For each branch

The posterior

$$p(\mathbf{f}|\mathbf{X}_k, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{f})p(\mathbf{f}|\mathbf{X}_k)}{p(\mathbf{y}_k|\mathbf{X}_k)}$$

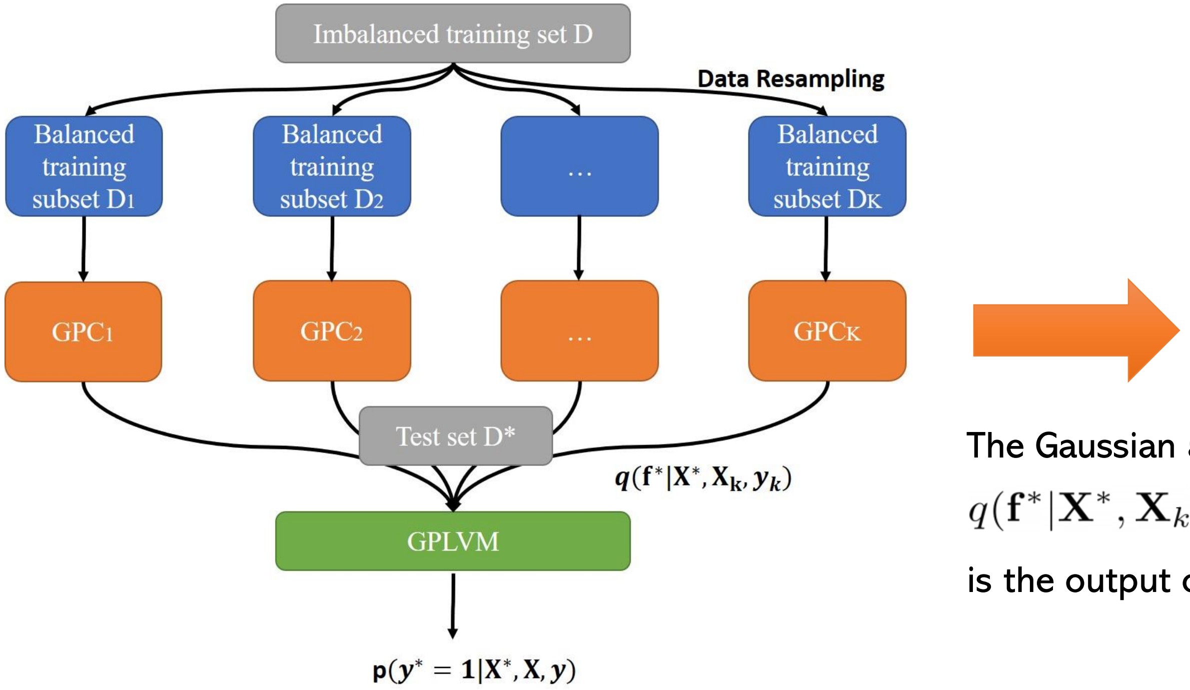
The predictive distribution

$$p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \int p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{f})p(\mathbf{f}|\mathbf{X}_k, \mathbf{y}_k)d\mathbf{f}$$

The Gaussian approximation

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

GPLVM for Ensemble Classification



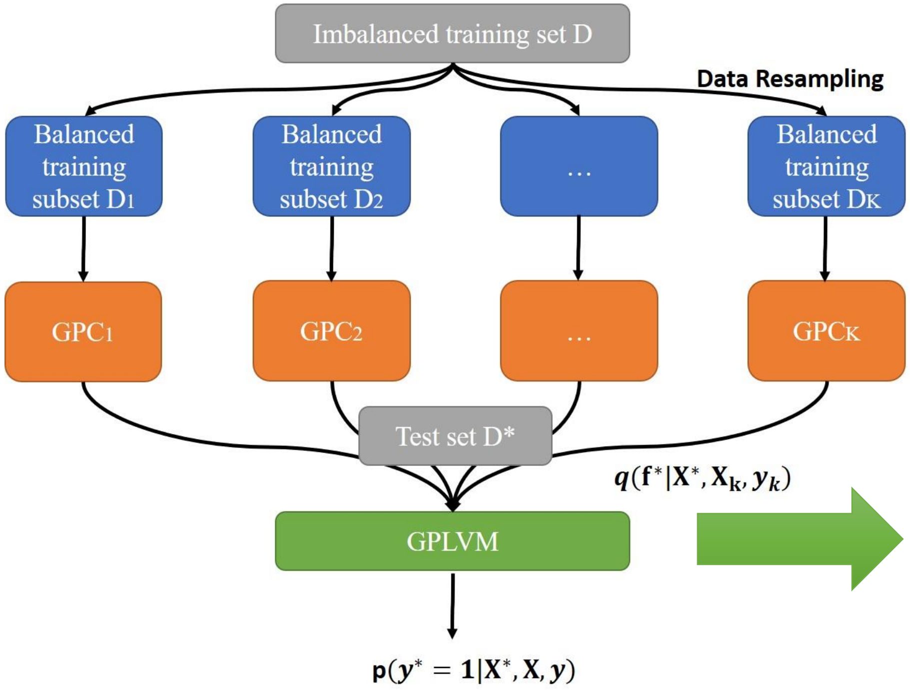
The Gaussian approximation

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

is the output of each Gaussian process classifier.

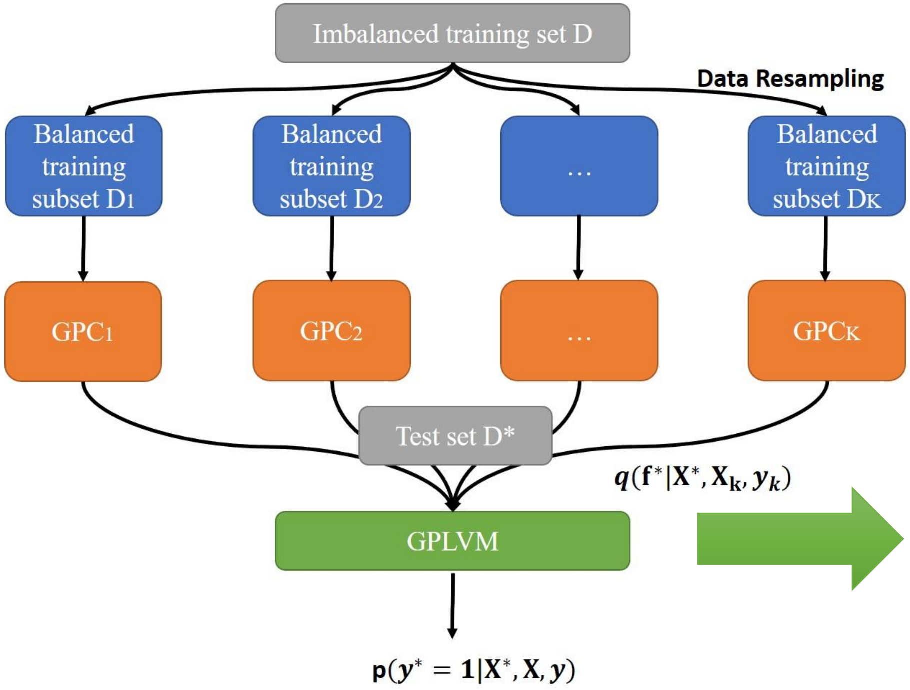
Stony Brook University

GPLVM for Ensemble Classification



Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

GPLVM for Ensemble Classification

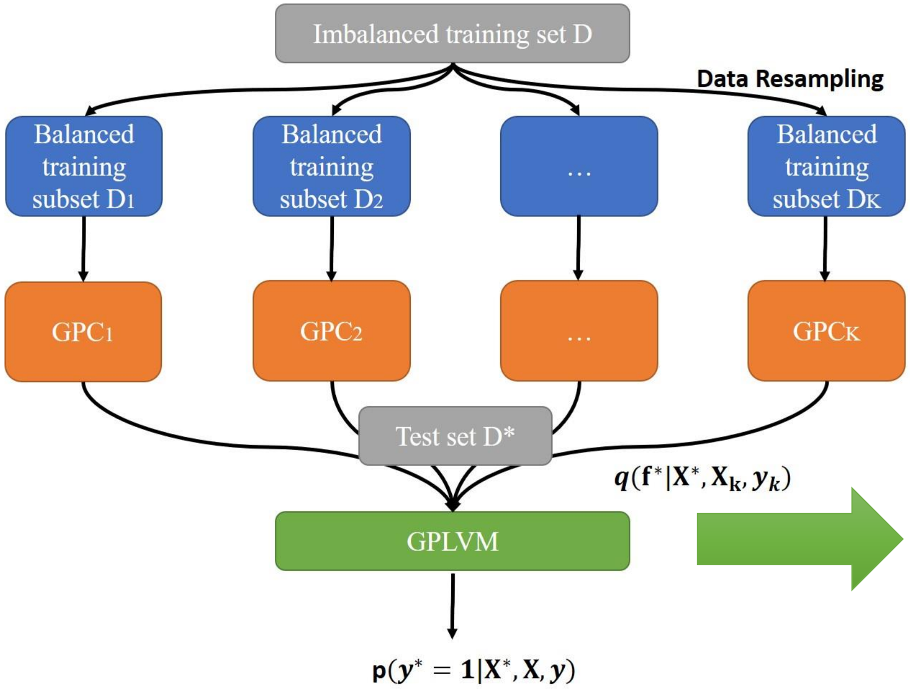


Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

GPLVM for Ensemble Classification



Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

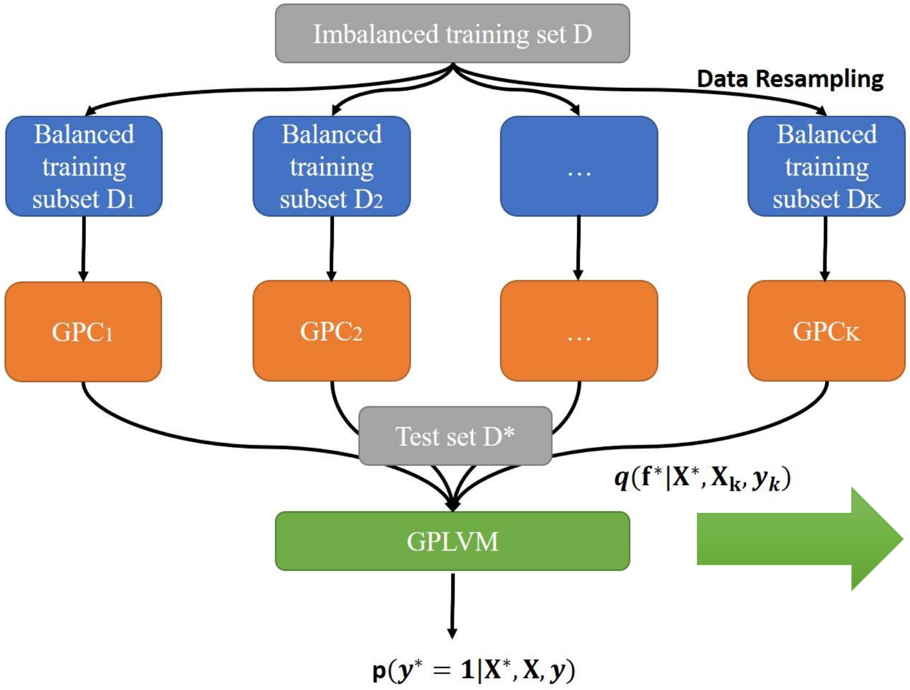
We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

GPLVM for Ensemble Classification



Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

E contains i.i.d. zero-mean Gaussian noises

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

GPLVM for Ensemble Classification

Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

E contains i.i.d. zero-mean Gaussian noises

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

GPLVM for Ensemble Classification

Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

E contains i.i.d. zero-mean Gaussian noises

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$
$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*) p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*) p(\mathbf{v}_k|\mathbf{f}^*) d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

GPLVM for Ensemble Classification

Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

E contains i.i.d. zero-mean Gaussian noises

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$
$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*) p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*) p(\mathbf{v}_k|\mathbf{f}^*) d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

GPLVM for Ensemble Classification

Let $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_K^*] \in \mathbb{R}^{n^* \times K}$ be an observation matrix
and $\mathbf{f}^* \in \mathbb{R}^{n^*}$ be an unknown true test latent vector.

We have a nonlinear mapping

$$\mathbf{F}^* = \mathbf{G}(\mathbf{f}^*) + \mathbf{E}$$

Function G(•) defines K independent GPs

$$\mathbf{f}_k^* = \mathbf{g}_k(\mathbf{f}^*) \sim \mathcal{GP}(\mathbf{m}(\mathbf{f}^*), \mathbf{K}(\mathbf{f}^*, \mathbf{f}^*))$$

E contains i.i.d. zero-mean Gaussian noises

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$
$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*) p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*) p(\mathbf{v}_k|\mathbf{f}^*) d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

The likelihood is a product of Gaussian distributions

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{f}_k^*|\mathbf{0}, \mathbf{\Sigma}_k)$$
$$\mathbf{\Sigma}_k = \mathbf{K}_k + \mathbf{\Sigma}_{\mathbf{f}^*,k} + \sigma^2 \mathbf{I}$$

FAR BEYOND

GPLVM for Ensemble Classification

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$

$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*)p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*)p(\mathbf{v}_k|\mathbf{f}^*)d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

The likelihood is a product of Gaussian distributions

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{f}_k^*|\mathbf{0}, \mathbf{\Sigma}_k)$$

$$\mathbf{\Sigma}_k = \mathbf{K}_k + \mathbf{\Sigma}_{\mathbf{f}^*,k} + \sigma^2\mathbf{I}$$

GPLVM for Ensemble Classification

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$

$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*)p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*)p(\mathbf{v}_k|\mathbf{f}^*)d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \boldsymbol{\Sigma}_{\mathbf{f}^*,k})$$

The likelihood is a product of Gaussian distributions

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{f}_k^*|\mathbf{0}, \boldsymbol{\Sigma}_k)$$

$$\boldsymbol{\Sigma}_k = \mathbf{K}_k + \boldsymbol{\Sigma}_{\mathbf{f}^*,k} + \sigma^2\mathbf{I}$$

$\mathbf{K_k}$ is the covariance matrix computed by evaluating the kernel of the kth GP.

Stony Brook University

GPLVM for Ensemble Classification

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$

$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*) p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*) p(\mathbf{v}_k|\mathbf{f}^*) d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

The likelihood is a product of Gaussian distributions

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{f}_k^*|\mathbf{0}, \mathbf{\Sigma}_k)$$

$$\mathbf{\Sigma}_k = \mathbf{K}_k + \mathbf{\Sigma}_{\mathbf{f}^*,k} + \sigma^2\mathbf{I}$$

$\mathbf{K_k}$ is the covariance matrix computed by evaluating the kernel of the kth GP.

If the prior $$p(\mathbf{f}^*) = \prod_{i=1}^{n^*} \mathcal{N}(0, 1)$$

FAR BEYOND

GPLVM for Ensemble Classification

Then, the likelihood function is

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} p(\mathbf{f}_k^*|\mathbf{f}^*)$$

$$= \prod_{k=1}^{K} \iint p(\mathbf{f}_k^*|\bar{\mathbf{f}}_k^*)p(\bar{\mathbf{f}}_k^*|\mathbf{v}_k, \mathbf{f}^*)p(\mathbf{v}_k|\mathbf{f}^*)d\bar{\mathbf{f}}_k^* d\mathbf{v}_k$$

Considering the outputs from GPCs

$$q(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}_k, \mathbf{y}_k) = \mathcal{N}(\mathbf{f}^*|\bar{\mathbf{f}}_k^*, \mathbf{\Sigma}_{\mathbf{f}^*,k})$$

The likelihood is a product of Gaussian distributions

$$p(\mathbf{F}^*|\mathbf{f}^*) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{f}_k^*|\mathbf{0}, \mathbf{\Sigma}_k)$$

$$\mathbf{\Sigma}_k = \mathbf{K}_k + \mathbf{\Sigma}_{\mathbf{f}^*,k} + \sigma^2\mathbf{I}$$

$\mathbf{K_k}$ is the covariance matrix computed by evaluating the kernel of the kth GP.

If the prior $p(\mathbf{f}^*) = \prod_{i=1}^{n^*} \mathcal{N}(0, 1)$

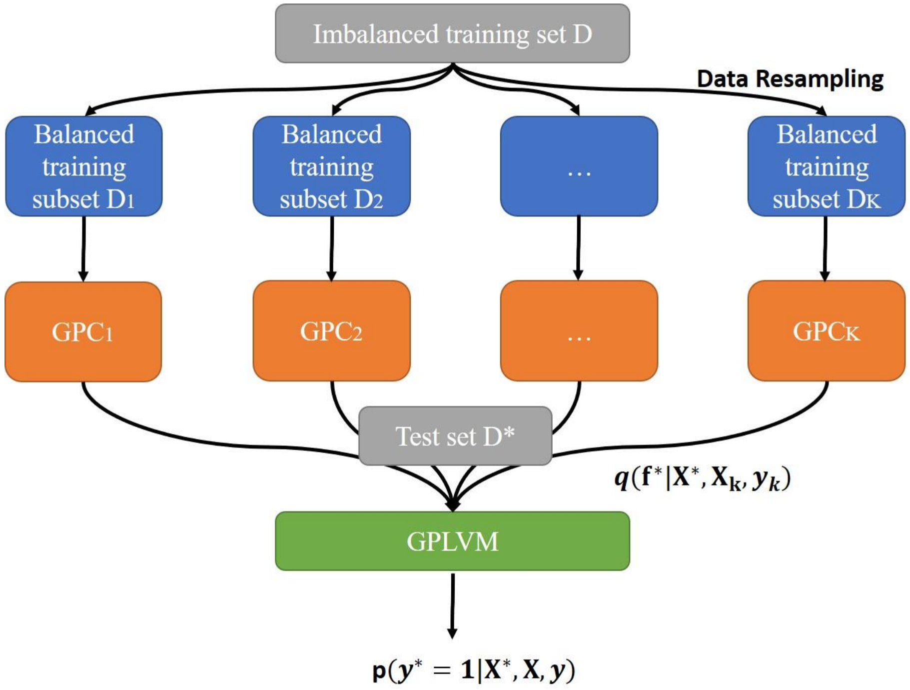
the log of the posterior

$$\log\, p(\mathbf{f}^*|\mathbf{F}^*) \propto -\frac{(K+1)n^*}{2}\log(2\pi) - \frac{1}{2}\mathrm{tr}(\mathbf{f}^*\mathbf{f}^{*T})$$

$$-\frac{1}{2}\sum_{i=1}^{n^*}\left(\log|\Sigma_k| + \bar{\mathbf{f}}_k^{*T}\Sigma_k^{-1}\bar{\mathbf{f}}_k^*\right)$$

Stony Brook University

GPLVM for Ensemble Classification



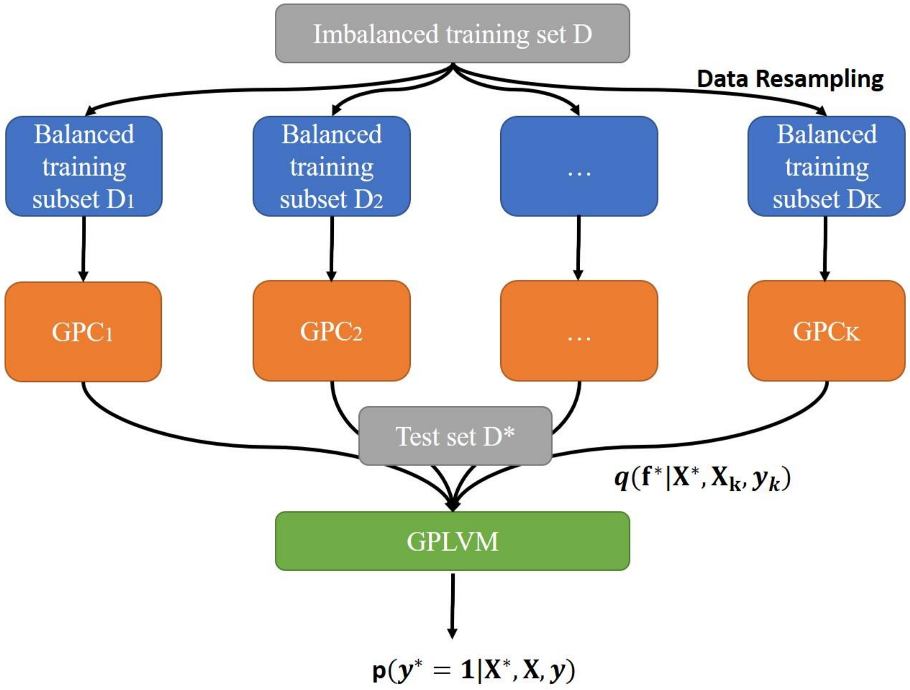
the log of the posterior

$$\log p(\mathbf{f}^*|\mathbf{F}^*) \propto -\frac{(K+1)n^*}{2}\log(2\pi) - \frac{1}{2}\mathrm{tr}(\mathbf{f}^*\mathbf{f}^{*T}) - \frac{1}{2}\sum_{i=1}^{n^*}\left(\log|\Sigma_k| + \bar{\mathbf{f}}_k^{*T}\Sigma_k^{-1}\bar{\mathbf{f}}_k^*\right)$$

FAR BEYOND

Stony Brook University

GPLVM for Ensemble Classification



the log of the posterior

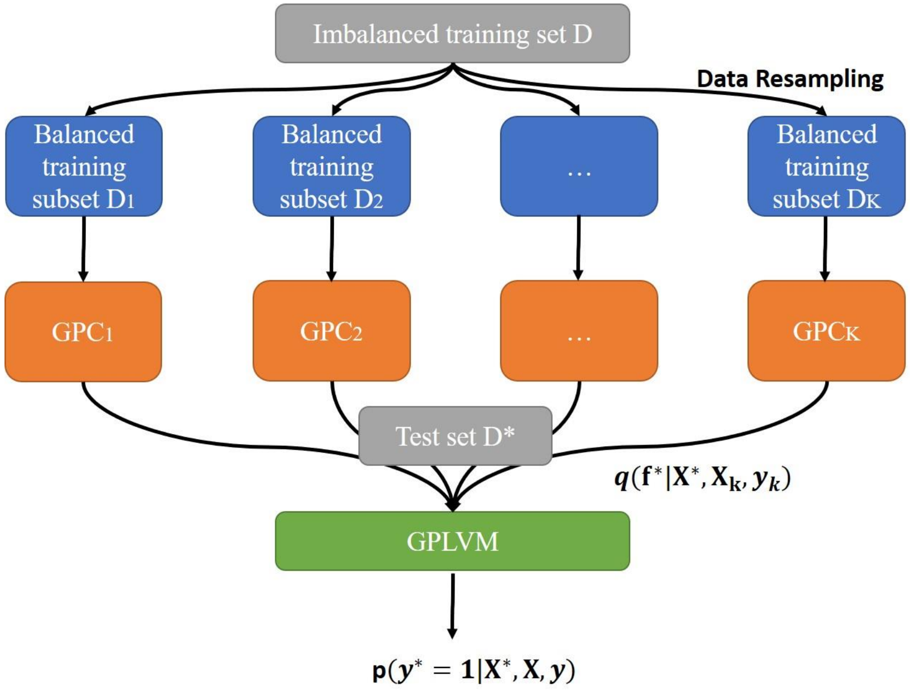
$$\log p(\mathbf{f}^*|\mathbf{F}^*) \propto -\frac{(K+1)n^*}{2}\log(2\pi) - \frac{1}{2}\mathrm{tr}(\mathbf{f}^*\mathbf{f}^{*T}) - \frac{1}{2}\sum_{i=1}^{n^*}\left(\log|\Sigma_k| + \bar{\mathbf{f}}_k^{*T}\Sigma_k^{-1}\bar{\mathbf{f}}_k^*\right)$$

The MAP estimation is

$$\hat{\mathbf{f}}^*_{\mathrm{MAP}} = \arg\max_{\mathbf{f}^*,\boldsymbol{\theta}} \log p(\mathbf{f}^*|\mathbf{F}^*)$$

FAR BEYOND

GPLVM for Ensemble Classification



the log of the posterior

$$\log p(\mathbf{f}^*|\mathbf{F}^*) \propto -\frac{(K+1)n^*}{2}\log(2\pi) - \frac{1}{2}\mathrm{tr}(\mathbf{f}^*\mathbf{f}^{*T}) - \frac{1}{2}\sum_{i=1}^{n^*}\left(\log|\Sigma_k| + \bar{\mathbf{f}}_k^{*T}\Sigma_k^{-1}\bar{\mathbf{f}}_k^*\right)$$

The MAP estimation is

$$\hat{\mathbf{f}}^*_{\mathrm{MAP}} = \arg\max_{\mathbf{f}^*, \boldsymbol{\theta}} \log p(\mathbf{f}^*|\mathbf{F}^*)$$

The probabilistic output is

$$p(\mathbf{y}^* = \mathbf{1}|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \phi(\hat{\mathbf{f}}^*_{\mathrm{MAP}})$$

GPLVM for Ensemble Classification

The MAP estimation is

$$\hat{\mathbf{f}}^*_{\text{MAP}} = \arg\max_{\mathbf{f}^*, \boldsymbol{\theta}} \log p(\mathbf{f}^* | \mathbf{F}^*)$$

The probabilistic output is

$$p(\mathbf{y}^* = \mathbf{1} | \mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \phi(\hat{\mathbf{f}}^*_{\text{MAP}})$$

GPLVM for Ensemble Classification

The MAP estimation is

$$\hat{\mathbf{f}}^*_{\text{MAP}} = \underset{\mathbf{f}^*,\boldsymbol{\theta}}{\arg\max} \log p(\mathbf{f}^*|\mathbf{F}^*)$$

The probabilistic output is

$$p(\mathbf{y}^* = \mathbf{1}|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \phi(\hat{\mathbf{f}}^*_{\text{MAP}})$$

We used the logistic function:

$$p(y_i|f_i) = \phi(y_i f_i) = \frac{1}{1 + \exp(-y_i f_i)}$$

GPLVM for Ensemble Classification

The MAP estimation is

$$\hat{\mathbf{f}}^*_{\text{MAP}} = \arg\max_{\mathbf{f}^*, \boldsymbol{\theta}} \log p(\mathbf{f}^* | \mathbf{F}^*)$$

The probabilistic output is

$$p(\mathbf{y}^* = \mathbf{1} | \mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \phi(\hat{\mathbf{f}}^*_{\text{MAP}})$$

We used the logistic function:

$$p(y_i | f_i) = \phi(y_i f_i) = \frac{1}{1 + \exp(-y_i f_i)}$$

We chose the popular radial basis function (RBF) as the kernel of GP models.

$$k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\Big(- \frac{||\mathbf{x} - \mathbf{x}'||^2}{2l^2} \Big)$$

And the automatic relevance determination (ARD) is applied for dimentionality reduction.

$$k_{\text{RBF-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\Big(- \frac{1}{2} \sum_{d=1}^{D} \frac{(x_d - x_d')^2}{l_d^2} \Big)$$

Experiments

Synthetic Binary Classification

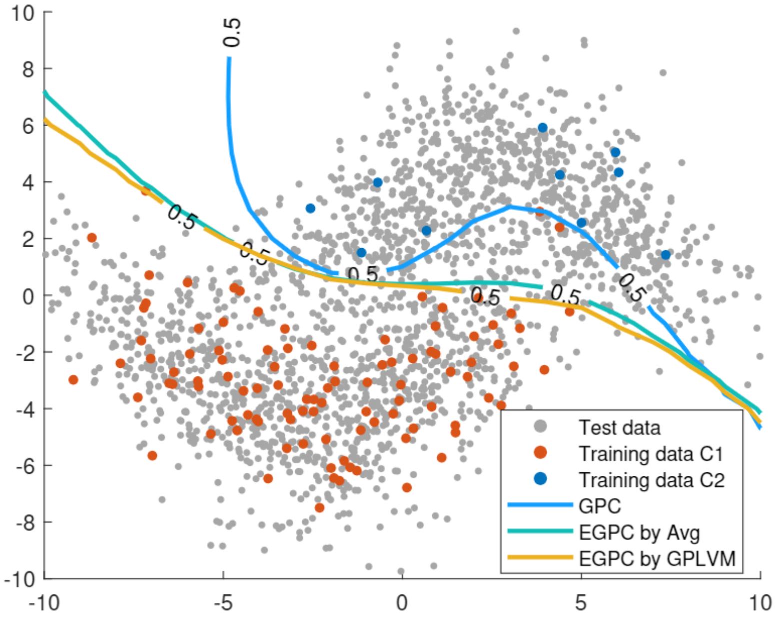
We generated a two-moon dataset centered at (2.5,3) and (-2.5,-3).

majority class: $n_{c_1} = 100$

minority class: $n_{c_2} = 10$

test set: $n^* = 2000$

# of branches: K=10
# of data in training subset: 20



Stony Brook University

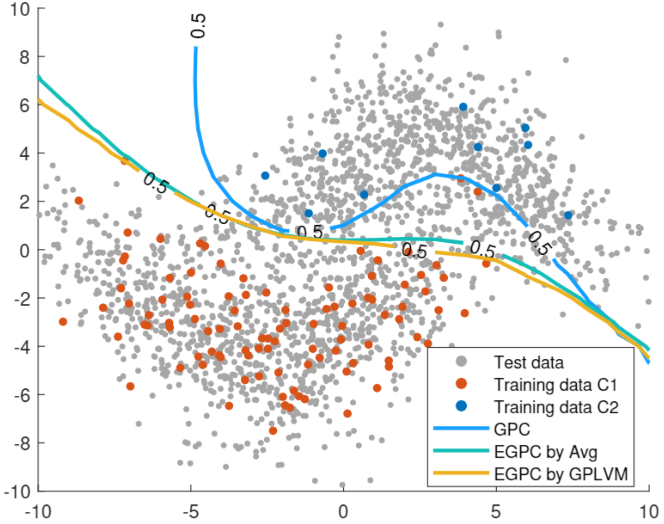
Experiments

Synthetic Binary Classification

EnGPC-GPLVM: newly proposed method

GPC: only using a GPC-based model on imbalanced dataset

EnGPC-Avg: ensemble of GPCs whose outputs are averaged



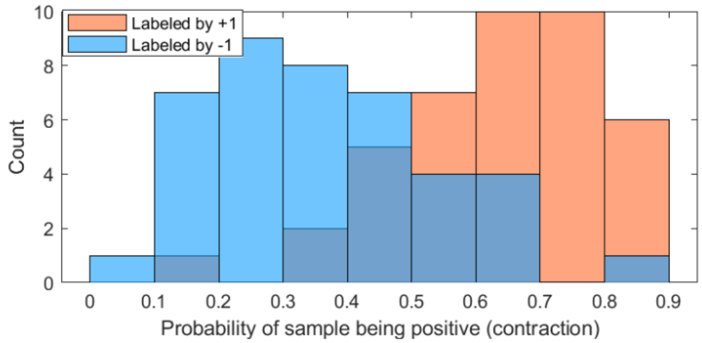
Methods	TPR	FPR	TNR	FNR	ACC	F-score
GPC	**0.9802**	0.2341	0.7659	**0.0115**	0.8702	0.8895
EnGPC-Avg	0.9408	0.0899	0.9190	0.0592	0.9200	0.9230
EnGPC-GPLVM	0.9331	**0.0638**	**0.9362**	0.0659	**0.9346**	**0.9347**

FAR BEYOND

Experiments

Test with real-world dataset

Based on the work in [1], we have a uterine contraction dataset annotated by experts.
Training set: 233 positive samples
46 negative samples
Test set: 41 samples per class



Methods	TPR	FPR	TNR	FNR	ACC	F-score
GPC	**0.9012**	0.3171	0.6829	**0.1488**	0.7171	0.7387
EnGPC-Avg	0.7561	0.1950	0.8049	0.2439	0.7805	0.7850
EnGPC-GPLVM	0.8049	**0.1195**	**0.8293**	0.1951	**0.8171**	**0.8148**

[1] L. Yang, C. Heiselman, J. G. Quirk and P. M. Djurić, "Identification of uterine contractions by an ensemble of Gaussian processes", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP-2021

Conclusions

- We addressed the problem of binary classification with imbalanced dataset.
- An ensemble of Gaussian process classifiers with the Gaussian process latent variable model as a decision maker, is proposed.
- Experiments using both synthetic and real-world data show promise of the proposed approach.

Thank you very much for your attention!

Contacts

*{liu.yang.2, petar.djuric}@stonybrook.edu

†{cassandra.heiselman, j.gerald.quirk}@stonybrookmedicine.edu