Hierarchical GPs for Outlier Detection

Felix Jimenez[1]

[1]National Institute of Standards and Technology

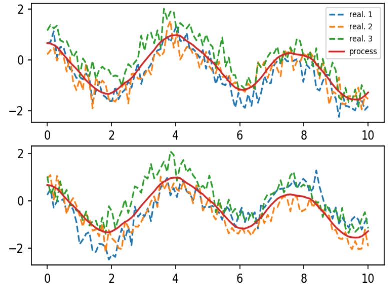
JSM, 2020

Outline

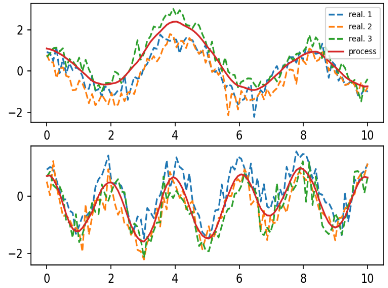
Motivating Problem

The Problem

Given two sets of time series we would like to know if they come from the same generative process.



Same process.



Different Processes.

Gaussian Processes Overview

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution (1).

Hierarchical Gaussian Processes (2)

We would like to model the dependence between multiple time series and one way we can do this is by assuming the following model:

$$g(t) \sim \mathcal{GP}(0, k_g(t, t')) \tag{1}$$
$$f_n(t) \sim \mathcal{GP}(g(t), k_f(t, t')) \tag{2}$$
$$y_n(t) \sim \mathcal{N}(f_n(t), \sigma) \tag{3}$$

Here n represents the repeat measurements of our true underlying process $g(t)$, $f_n(t)$ represents realizations of that process and $y_n(t)$ represents noisy observations of those realizations.

To represent our model in a format we are more familiar with consider stacking the repeat measurements on top of one another.

Hierarchical Covariance Structure (2)

By stacking the measurements we can write the log marginal-likelihood in a compact fashion:

$$log\ p(Y|\theta) = -\frac{D}{2} log 2\pi - \frac{1}{2} log|K_{t,t} + \sigma I| - \frac{1}{2} Y^T (K_{t,t} + \sigma I)^{-1} Y \quad (4)$$

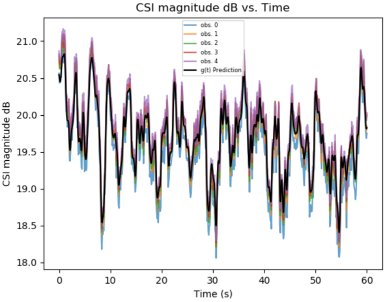
We can now populate the covariance matrix K using the following covariance relationships. (2):

$$cov(y_n(t), f_{n'}(t)) = \begin{cases} k_f(t) + k_g(t) & \text{if } n = n' \\ k_g(t) & \text{if } n \neq n' \end{cases} \quad (5)$$

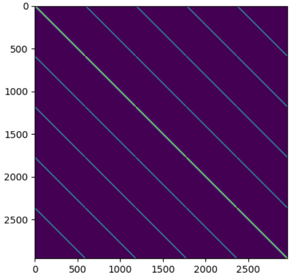
$$cov(y_n(t), g_{n'}(t)) = k_g(t, t) \quad (6)$$

We can choose any valid covariance functions for k_f and k_g, and then simply optimize over our hyper-parameters to fit our data.

Example of Hierarhical Time Series



Multiple Channel State Information (CSI) estimates combined to estimate underlying channel state.



Estimated covariance structure between the individual CSI estimates.

The Problem We Are Working on

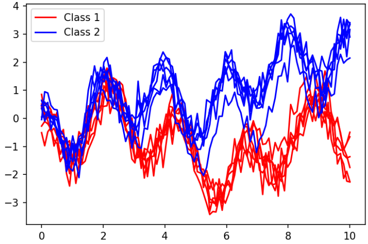
The question we want to answer:

- Can we accurately determine when two sets of time series come from different generative processes.

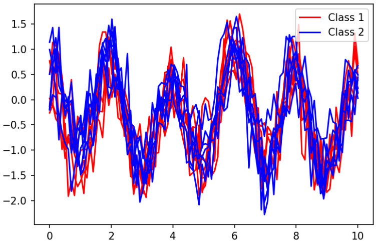
Identifying Outliers

Two approaches

- MSE on hold out data.
- Comparing estimates of underlying process.



What our outlier data looks like.



What our no outlier data looks like.

MSE on Hold Out Set

Hold out last 10% of each of the 12 time series and then use two competing models to predict hold out data.

Model 1

$$
\begin{aligned}
g_1 &\sim \mathcal{GP}(0, k_{g_1}(t, t')) \\
f_{i1} &\sim \mathcal{GP}(0, k_{f_1}(t, t')) \\
i &= 1, ..., N_1
\end{aligned}
$$

$$
\begin{aligned}
g_2 &\sim \mathcal{GP}(0, k_{g_2}(t, t')) \\
f_{i2} &\sim \mathcal{GP}(0, k_{f_2}(t, t')) \\
i &= N_1 + 1, ..., N
\end{aligned}
$$

Model 2

$$
\begin{aligned}
g &\sim \mathcal{GP}(0, k_g(t, t')) \\
f_i &\sim \mathcal{GP}(0, k_f(t, t')) \\
i &= 1, ..., N
\end{aligned}
$$

Results for MSE on Hold Out Set

Evaluating Model

- Hold out last 10% of each of the 12 time series.
- Fit the two models.
- Compute the MSE for each model on the hold out data.

Results

Outlier Data	No Outlier
MSE(M1)/MSE(M2) = .922	MSE(M1)/MSE(M2) = 1.12

Comparing Estimates of Processes

Basic Idea

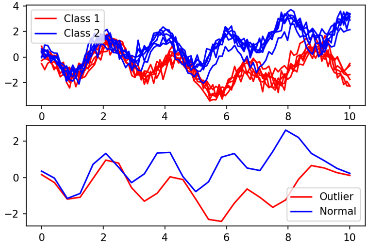
- Fit separate hierarchical models to both sets of time series.
- Compute estimates of the underlying process for both models.
- Compute distance between these estimates.
- Compare computed distance to a null distribution.

Null Distribution

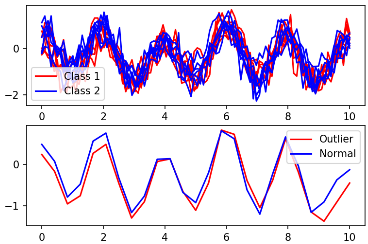
$$g(t) \sim \mathcal{GP}(0, k_g(t, t'))$$

Simulate a realization of $g(t)$ using the covariance matrix fit on the data. Next compute the distance between the $g(t)$ estimate and this simulated $g(t)$. Repeat this process 1000 times.

Results for Process Comparison



Outlier data.

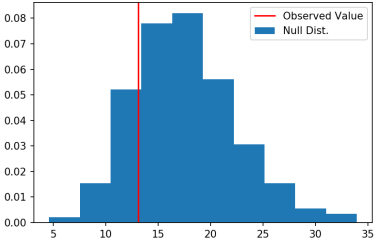


No outlier data.

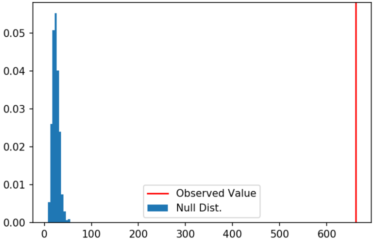
Results for Comparing Estimates of Processes

Results

Below we have the null distribution and the observed distance for no outlier data (left) and outlier data (right)



No outlier data.



Outlier data.

Summary

Topics Covered

- Hierarhical Gaussian processes.
- Outlier detection schemes.
- Results on simulated Data.

Next Steps

- Use method on other types of data.
- Develope solution for finding a single outlier rather than a group of outliers.

Contact

felix-jimenez.com

[1] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[2] J. Hensman, N. D. Lawrence, and M. Rattray, "Hierarchical bayesian modelling of gene expression time series across irregularly sampled replicates and clusters," *BMC bioinformatics*, vol. 14, no. 1, p. 252, 2013.

Supplementary Slides

Distance

$$D(g_1, g_2) = (\mu_1 - \mu_2)\Sigma^{-1}(\mu_1 - \mu_2)^T$$
$$\Sigma = \frac{\Sigma_1 + \Sigma_2}{2}$$

Simulating Data

In Class:

$$\text{g}_t = \mathcal{N}(cos(\pi t), K_g)$$
$$\text{y}_t = g_t + \sigma$$

Out Class:

$$\text{g}_t = \mathcal{N}(cos(\pi t * 0.9), K_g)$$
$$\text{y}_t = g_t + \sigma$$