Gaussian Process to Model Human Motion

Upayan Mathkari
upayan@cs.utexas.edu

The University of Texas at Austin
Dept. of Mechanical Engineering

1. Introduction

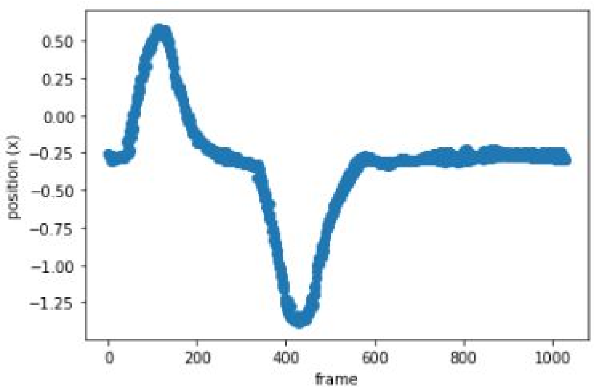
Gaussian processes provide a non-parametric method to perform regression, a tool used to make predictions about data based on prior knowledge. This method assumes that the data, evidence in the Bayesian inference model, are sampled from a multivariate Gaussian distribution. Then, the desired interpolation and extrapolation datapoints form conditional slices of the multivariate distribution and by the properties of Gaussians, are Gaussians themselves. Mean values and confidence intervals for these datapoints can then be easily gleaned if a covariance matrix, the prior, is known. The kernel function is used to determine this covariance matrix and can be optimized by tuning its hyperparameters.

In this project, we were given motion data tracking positions of various parts of a body in three-dimensional coordinates over time. The data was collected as the subjects traced curves and our goal is to model this motion using a Gaussian process. Due to the complexity of the data, we focus on a single coordinate for a single marker for one of the subjects: the x-component of the finger tracing out a circle.

2. Methods

Preprocessing

The subject analyzed, AG, performed five trials of the same motion. To best capture the information captured in all five trials while keeping the computational costs of the experiment low, the data was processed such that positional data was randomly sampled from one of the five trials for each of the 1030 frames of time that position was recorded.



Kernel Selection

We utilized a radial basis function kernel:

$$k(t, t^*) = \exp(\sigma_f) \exp(-0.5 \exp(\sigma_l) |t - t^*|^2) + \exp(\sigma_n) \boldsymbol{I}$$

Where t and t* are time vectors, and σ_f, σ_l, and σ_n are hyperparameters representing the covariances of the processes at a given input, dependence of ouputs on the separation of their inputs and assumed noise in the data. As a result, for a time vector of length n, the kernel function yields a n x n covariance matrix.

Hyperparameter Determination

The hyperparameters can be learned using a log-likelihood function of the Gaussian process model, and maximizing with respect to the hyperparameters given a training data set. This maximization typically requires using gradient descent methods on the following function:

$$\log P(y|f, \theta) = 0.5 f^T K^{-1} f - 0.5 \log|K| - constant$$

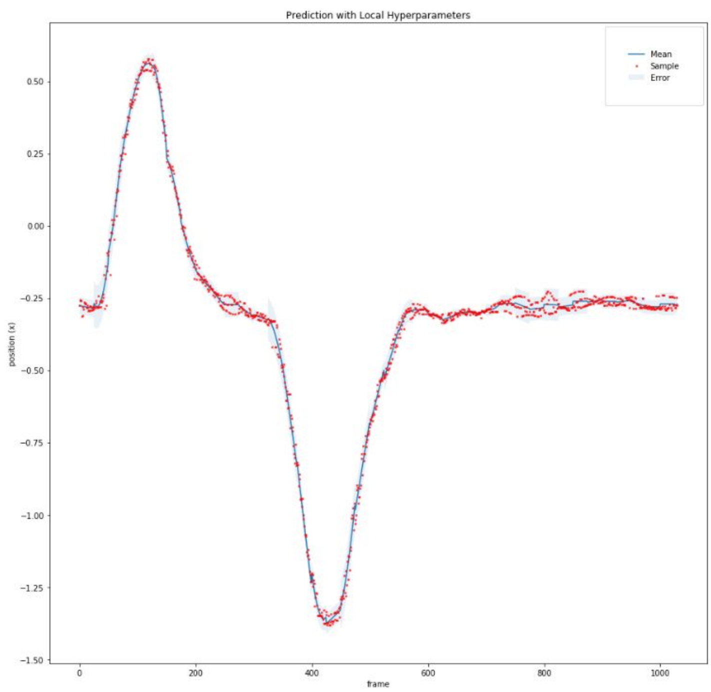
For this project, the gpr() package from sci-kit learn was used to perform this gradient descent to determine the optimal hyperparameters globally first and then, using a sliding window, locally. The performance of the global and local hyperparameters was then compared by evaluating which combination resulted in a higher total log mean likelihood value and then visually inspecting the resulting plots to determine which combination yielded a better fit.

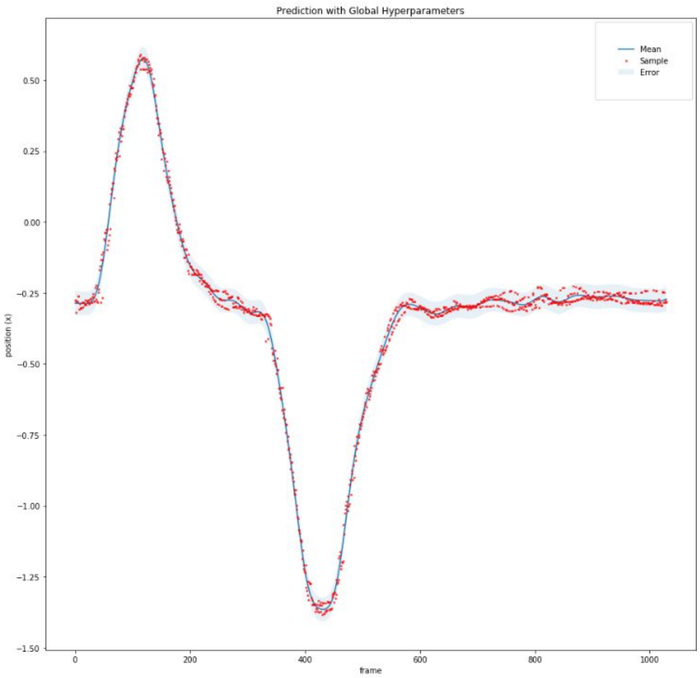
Prediction of Motion Data

As referenced in the introduction, interpolation and extrapolation points represent conditional slices of the multivariate Gaussian distribution. Once the optimal hyperparameters are determined, the gpr.predict() function can be used to determine the mean and standard deviation for each datapoint. The 95% confidence interval around each predicted mean point as well as the mean is then plotted to represent the mean function and the error associated for the process. The same process can be used with local hyperparameters.

3. **Results**

Prediction Plots

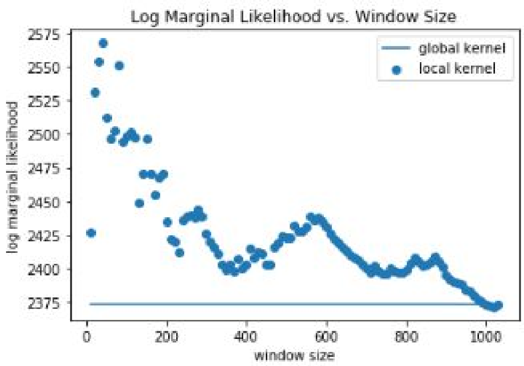




The prediction of motion data using global and local parameters (window size=40) are plotted above, respectively. Qualitatively, the error region, shaded in blue, appears to be much smaller for the plot with local hyperparameters suggesting that the local kernels yielded a better fit.

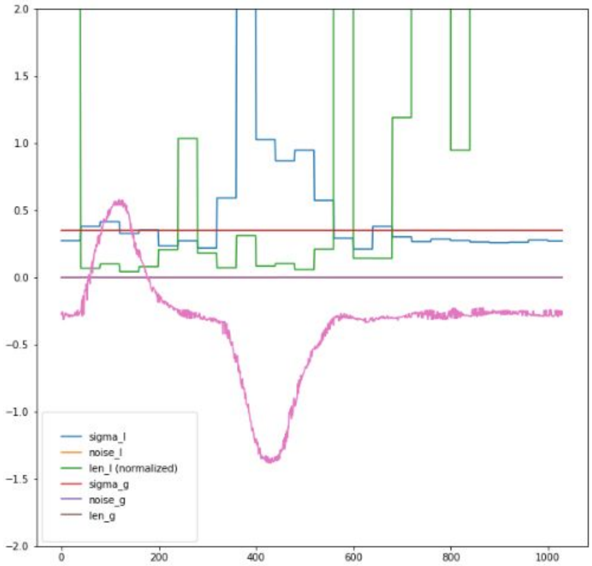
Log Marginal Likelihood

To determine whether local kernels provided a better fit quantitatively, the total log marginal likelihood was determined for window sizes incrementally varying from 10 to 1030, where a window size of 1030 represents a global kernel fit.



As the figure above demonstrates, for almost all window sizes smaller than ~1030 the local kernels outperform the global kernel in accuracy. The peak value (log marginal likelihood = 2567) was used to determine an optimal window size of 40 which was used for obtaining the remaining figures.

Hyperparameters vs. Time



The above figure demonstrates how the local hyperparameter vary across frames when using a sliding window of size 40. The local hyperparameters appear to fluctuated a lot across the sinusoidal portion of the motion but then appear to level off during the flatter portion indicating that this section can be modelled by more constant parameters.

4. **Summary**

In this project, Gaussian process regression was used to fit a function to a dataset modelling a human finger tracing out a curve in three-dimensional space by maximizing the probability of test points given the distribution of training points. The regression was implemented using a global kernel with constant hyperparameters and local kernels with hyperparameters varying based on user-determined window size. By visual inspection and comparison of the log marginal likelihood for both methods, it is evident that use of local kernels yields a higher accuracy fit.