Introduction to Gaussian Processes SPA DRP Autumn 2024

Student: Arnav Palkhiwala

Mentor: Wolfgang Brightenburg

I. Introduction

Supervised learning is when machine learning models utilize labeled datasets to represent relationships between variables. Gaussian Processes is a non-parametric (not dependent on any distribution) probabilistic approach to finding a function. It gives function predictions to fit the training data along with uncertainty measures for each prediction, using the ideas of the Gaussian Distribution applied to functions. Gaussian Processes can be used for linear and multiple regression as well as classification, though this paper will focus on regression. In mathematical terms, Gaussian Processes model data using multivariate normal distribution samples with a covariance function.

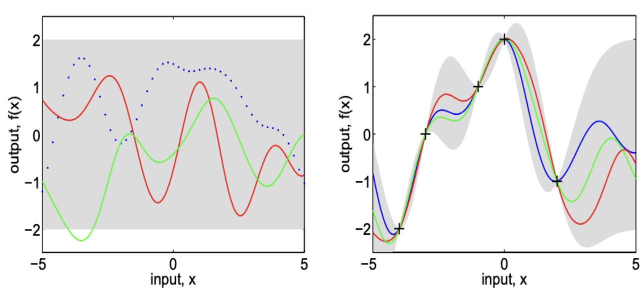
II. Strengths and Weaknesses of Gaussian Processes.

Strengths: They are flexible in the ability to model non-linear relationships very precisely. The uncertainty percentage from the predictive distribution over all possible outputs provides a unique perspective on the results, and this method looks at a finite amount of points without requiring a significant amount of data. **Weaknesses**: In some situations, simpler models such as linear regression can be more effective and can be computationally expensive, since the inversion of an nxn covariance matrix has a runtime complexity of $O(N^3)$. Furthermore, the accuracy of the results of Gaussian Processes are dependent on the kernel function selection.

III. Kernels

Kernels are used to measure the similarities between data points in different dimensions and are effective in representing nonlinear functions. They measure the impact of each input in the model, creating a dot product of two points as if they are in a higher dimension without actually being there, allowing for the higher dimension to be simulated without computational intensity. This is the differentiating factor.

IV. Space of Prior and Posterior

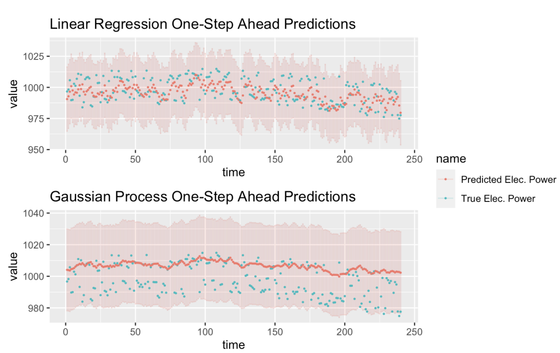


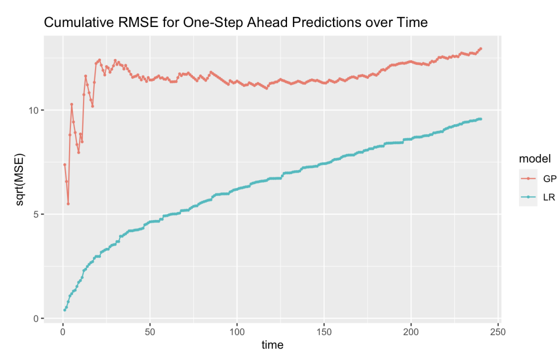
(a), prior (b), posterior

The prior is before Gaussian Process has been applied, our prediction of the true function (kernel chosen with this assumption). The posterior shows the Gaussian Process' accurate prediction of the true function, with the gray space around the function being the uncertainty measure to provide a margin for error against the prediction. Obtaining data and looking at the noisy function evaluations on those data points gives us a distribution of functions conditioned on our dataset. This results in the posterior mean, kernel regression, and posterior variance functions. (Reference: This visualization is from GP for Machine Learning Textbook)

V. Application of Gaussian Process

Here is an application of Gaussian Process of Electricity Production over Time when Power is Applied to a Gas Generator. We can see Gaussian Process' effectiveness as predictions show good results but also learn that in simpler data sets linear regression can be a better choice.





VI. References

Rasmussen Gaussian Process for Machine Learning Textbook, MIT Lectures on Neural Tangent Kernels, Murphy's Textbook