On Gaussian Processes for Regression

Anonymous Author(s)
Affiliation
Address
email

Abstract

Gaussian processes emerged in machine learning as a powerful tool for regression and classification that provides interpretability through kernel choice and uncertainty quantification. By leveraging properties of multivariate normal distributions and Bayes's rule, we may infer a probability distribution over possible functions when fitting a dataset. This Bayesian framework allows flexibility through choosing a covariance function as a prior belief about the dataset, which can provide further insight into the trends of the training data. We implement a multi-dimensional Gaussian process regressor and evaluate its performance on the Boston Housing dataset, which is comparable to those in the top 25 of the Kaggle competition. Furthermore, we perform optimization on the hyperparameters through maximum likelihood estimation, to remove the need for manual tuning of the hyperparameters.

1 Gaussian Random Variables

A random variable is a function that maps from an event space to a measurable space. The event space represents a set of all possible outcomes that the random variable may take, and the measurable space is a probability measure between 0 and 1 (inclusive). We say that a random variable X is normally distributed if the event space has a Gaussian probability distribution, fully characterized by two parameters: a mean μ and variance σ^2:

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

For a one-dimensional Gaussian random variable, we refer to its distribution as a univariate Gaussian distribution. A set of Gaussian random variables may be characterized jointly as a multivariate Gaussian distribution, with joint probability distribution fully characterized by a mean vector and a covariance matrix:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

where $\boldsymbol{\mu}$ is the mean vector, and Σ is the covariance matrix whose entries describe the covariance between each pair of random variables.

2 Gaussian Process

A random process is essentially a collection of random variables jointly characterized as a set or vector of random variables with a multivariate joint probability distribution. A Gaussian process $f(\boldsymbol{x})$ is defined as a a random process where each set of random variable in the random process is

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

has a multivariate Gaussian distribution. $f(\boldsymbol{x})$ is fully characterized by a mean function $m(\boldsymbol{x})$ and covariance function, or kernel $K(\boldsymbol{x}, \boldsymbol{x'})$:

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x'}))$$

Typically, the mean function is zero. The kernel is chosen based on some prior belief about the dataset; more on kernels is discussed in 3.1.

3 Regression

Suppose we observe training data $\boldsymbol{t}$ and choose covariance function κ. Then the mean function is given by

$$\begin{aligned} m(\boldsymbol{x}) &= C_{\boldsymbol{xt}}^{\top} C_{\boldsymbol{t}}^{-1} \boldsymbol{t} \\ K(\boldsymbol{x}, \boldsymbol{x'}) &= C_{\boldsymbol{xx'}} - C_{\boldsymbol{xt}}^{\top} C_{\boldsymbol{t}}^{-1} C_{\boldsymbol{xt}} \end{aligned}$$

where $C_{\boldsymbol{xt}} = \kappa(\boldsymbol{x}, \boldsymbol{t})$, $C_{\boldsymbol{t}} = \kappa(\boldsymbol{t}, \boldsymbol{t})$, and $C_{\boldsymbol{xx'}} = \kappa(\boldsymbol{x}, \boldsymbol{x'})$. Those interested in the derivation of the results are encouraged to consult section 6.4.2 of [1].

3.1 Kernels

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Note that the Reference section does not count towards the eight pages of content that are allowed.**

[1] C. M. Bishop, Pattern recognition and machine learning. New York: Springer, 2006.