Approximate Hyperparameter Marginalisation for Gaussian Processes

Anonymous Author(s)
Affiliation
Address
email

Abstract

Enter Abstract here

1 Introduction

2 Gaussian Processes

Gaussian processes (GPs) constitute a powerful method for performing Bayesian inference about functions using a limited set of observations [1]. A GP is defined as a distribution over the functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that the distribution over the possible function values on any finite set of $\mathcal{X}$ is multivariate Gaussian. Given some arbitrary size n dataset, the observations $\mathbf{y} = \{y_1, ..., y_n\}$ could be viewed as a single point sampled from a n-variate Gaussian distribution and can be partnered with a GP.

A GP is completely defined by its first and second moments: a mean function $\mu : \mathcal{X} \rightarrow \mathbb{R}$ which describes the overall trend of the function, and a symmetric positive semidefinite covariance function $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ which describes how function values are correlated as a function of their locations in the domain. Given a function $f : \mathcal{X} \rightarrow \mathbb{R}$ and a set of input points $\mathbf{x} \subset \mathcal{X}$, the Gaussian process prior distribution over the function values $\mathbf{f} = f(\mathbf{x})$ is given by:

$$
\begin{aligned}
p(\mathbf{f}|\mathbf{x}, \boldsymbol{\theta}) &= \mathcal{N}\left(\mathbf{f}; \mu(\mathbf{x}, \boldsymbol{\theta}), K(\mathbf{x}, \mathbf{x}, \boldsymbol{\theta})\right) && (1) \\
&= \frac{1}{\sqrt{\det 2\pi K_{\mathbf{f}}}} \exp\left(-\frac{1}{2}(\mathbf{f} - \mu_{\mathbf{f}})^\top K_{\mathbf{f}}^{-1}(\mathbf{f} - \mu_{\mathbf{f}})\right) && (2)
\end{aligned}
$$

where $\boldsymbol{\theta}$ is a vector containing any parameters required by μ and K, and which form the *hyperparameters* of the model. There exists a wide variety of mean and covariance functions which can be chosen in order to reflect any prior knowledge available about the function of interest.

3 Approximate Hyperparameter Marginalisation

3.1 Proof of Positive Semi-Definiteness

A sum of kernels is itself a kernel, which by definition fulfils the necessary condition of positive semi-definiteness. Therefore:

$$
K = \int k(\beta) p(\beta) d\beta \qquad (3)
$$

4 Experiments

5 Related Work

6 Conclusion

Acknowledgments

Do we have any? Aladdin / Orchid?

References

[1] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press Cambridge, MA, USA, 2006.