AIMS Course 1: Data, Estimation and Inference

Jake Levi

October 2022

1 Introduction

This lab report investigates the use of Gaussian Processes, a type of machine learning model motivated by Bayesian probability theory, for modelling a meteorological dataset called Sotonmet. In a Gaussian Process model, given a mean function μ, kernel function K, training input data x (represented as a vector), and prediction inputs x^*, the noisy training labels y (which we assume are noisy observations of unknown noiseless training labels f) and noiseless prediction labels f^* have a joint Gaussian distribution:

$$p\left(\begin{bmatrix} y \\ f^* \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} y \\ f^* \end{bmatrix} \middle| \begin{bmatrix} \mu(x) \\ \mu(x^*) \end{bmatrix}, \begin{bmatrix} K(x,x) + \sigma^2 I & K(x,x^*) \\ K(x,x^*)^T & K(x^*,x^*) \end{bmatrix}\right) \tag{1}$$

Where $K(x, x^*)$ is a matrix whose (i, j)th element is given by $K(x, x^*)_{i,j} = K(x_i, x_j^*)$. The predictive distribution $p(f^* \mid y)$ follows from the formula for the conditional distribution of a jointly Gaussian random variable [1]:

$$p(f^* \mid y) = \mathcal{N}\left(f^* \mid \mu^*, \Sigma^*\right) \tag{2}$$

$$\text{where} \qquad \mu^* = \mu(x^*) + K(x^*, x)\left(K(x,x) + \sigma^2 I\right)^{-1}(y - \mu(x)) \tag{3}$$

$$\Sigma^* = K(x^*, x^*) - K(x^*, x)\left(K(x,x) + \sigma^2 I\right)^{-1} K(x, x^*) \tag{4}$$

The log marginal likelihood (LML) of the noisy training labels y given training input data x (and also implicitly given any hyperparameters of the model) is given by:

$$\log\left(p(y \mid x)\right) = -\frac{1}{2}\log\left(\det\left(2\pi\left(K(x,x) + \sigma^2 I\right)\right)\right) - \frac{1}{2}(y - \mu(x))^T\left(K(x,x) + \sigma^2 I\right)^{-1}(y - \mu(x)) \tag{5}$$

This expression implies that maximising the LML encourages $\mu(x)$, $K(x, x)$ and σ to fit the data accurately and with calibrated uncertainty. Maximising the LML can also be motivated from a Bayesian perspective, which is discussed further in appendix B.

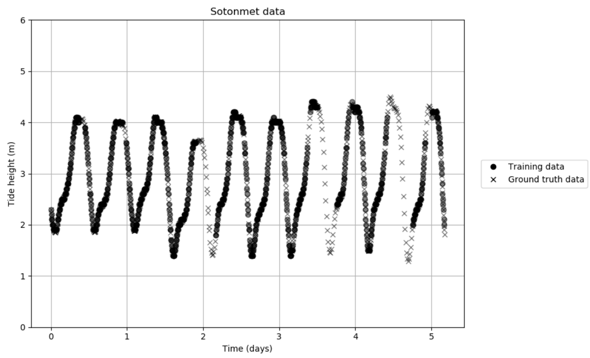
The focus in this coursework submission is on predicting the tide height given the time of day, for which the the training and ground truth data is shown in 1a, alongside some independent Gaussian Process predictions in 1b.

A Figures

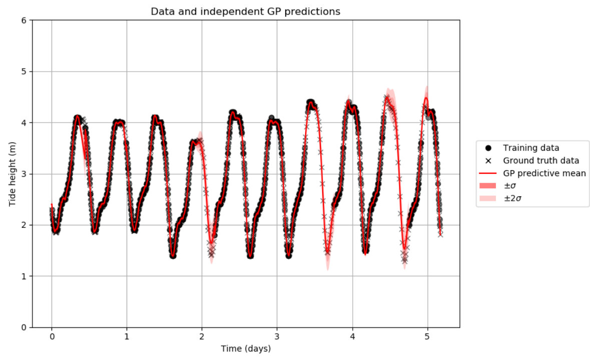
B Motivation for maximising the log marginal likelihood

References

[1] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.



(a) Training and ground truth data



(b) Independent Gaussian Process predictions

Figure 1: The Sotonmet dataset