4F13 Coursework 1 - Gaussian Processes

October 2024

1 Task A

```
meanfunc = []; hyp_init.mean = [];
covfunc = @covSEiso; hyp_init.cov = [−1 0];
likfunc = @likGauss; hyp_init.lik = 0;
hyp_opt = minimize(hyp_init, @gp, −100, @infGaussLik,
                   meanfunc, covfunc, likfunc, x, y);
[mu, s2] = gp(hyp_opt, @infGaussLik, meanfunc, covfunc, likfunc, x, y, xs);
```

Listing 1: Code to train hyper-parameters and generate the predictive distribution of a GP with squared exponential covariance

Our model has a GP with squared exponential (SE) covariance function and zero mean along with a gaussian noise. The SE covariance function has two hyper-parameters, λ which controls the characteristic length scale of the process and σ_f^2 which controls the variance in the output value of the GP. The gaussian noise has one hyper-parameter, σ_n^2, the variance of our measurement noise.

$$y = f(x) + \eta : \eta \sim \mathcal{N}(0, \sigma_n^2)$$
$$f = \mathcal{N}(0, k(x, x'))$$
$$k(x, x') = \sigma_f^2 \exp(-\frac{(x - x')^2}{2\lambda^2})$$

	λ	σ_f	σ_n	$\log(Z_{|\mathbf{y}})$
Initial	0.3679	1	1	-9.2910×10^1
Optimised	0.1282	0.8970	0.1178	-1.1899×10^1

Table 1: Initial and optimum hyper-parameter values when maximising log marginal likelihood of data

When we train our model hyper parameters to maximise log marginal likelihood we achieve the optimum values shown in Table 1. The covariance length scale, λ is similar to the length scale of the undulations in our data. The σ_f is much larger than σ_n this makes sense as most of the variation in our data can be explained by our GP instead of the measurement noise.

We plot the mean and 95% error bars of the predictive distribution (Equ.1) in Fig 1. We can see that the predictive error bars are smaller where we have data and return to the prior variance $(k(x, x) = \sigma_f^2)$ plus σ_n^2 when there is no nearby data. This is clear from the form of the predictive distribution covariance, the first term is the prior covariance, the second is the noise variance and the third reduces the variance based on the distance between the sample point and our data.

$$y_* \sim \mathcal{N}(m_{|\mathbf{y}}(x_*), k_{|\mathbf{y}}(x_*, x_*))$$
$$m_{|\mathbf{y}}(x) = m(x) + k(x, \mathbf{x})[k(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}(\mathbf{y} - m(\mathbf{x})) \tag{1}$$
$$k_{|\mathbf{y}}(x, x') = k(x, x') + \sigma_n^2 - k(x, \mathbf{x})[k(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}k(\mathbf{x}, x')$$
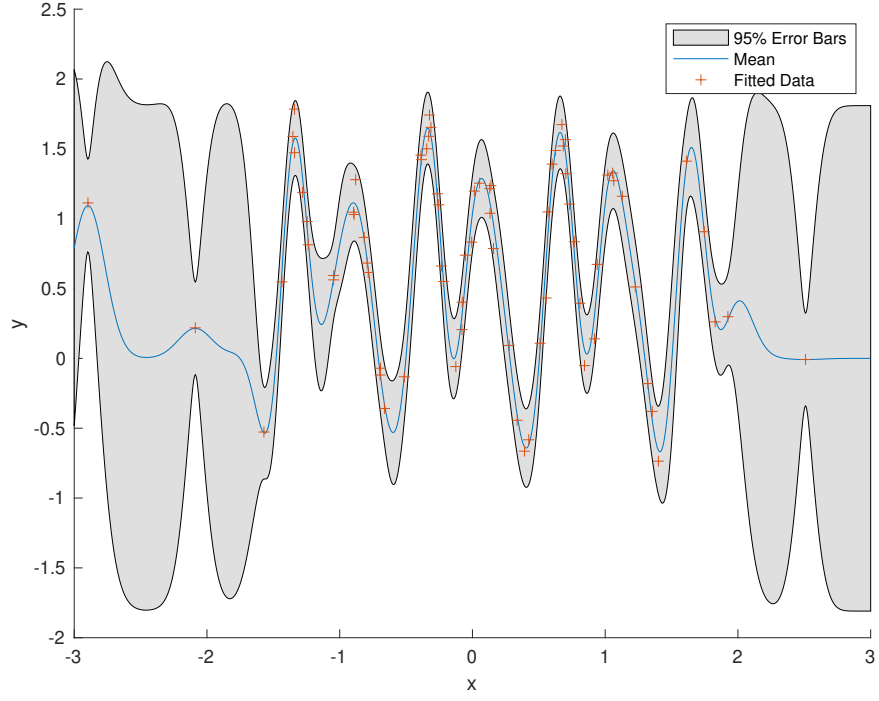
Figure 1: Mean and 95% prediction interval of GP with squared exponential covariance function, optimised hyper-parameters used given in Table 1

2   Task B