Appendix – GPR

A Gaussian process is defined as a collection of random variables that have a joint Gaussian distribution $f(\boldsymbol{x})$. Using Bayesian inference, we assume that the *prior* distribution can be described by a zero-mean Gaussian process with covariance function $k(\boldsymbol{x}, \boldsymbol{x})$:

$$\begin{bmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_n) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix}\right)$$

We then use data to train our model, in order to find the *posterior* distribution. In the training process, the functions that pass through the data points are given a higher likelihood and thus have a greater impact on the posterior distribution. Denoting the covariance matrix as $K(X, X)$ and letting $\boldsymbol{f}_*$ describe the function values at the data points, we get a joint distribution of measured values $\boldsymbol{f}$ and predicted values $\boldsymbol{f}_*$:

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right)$$

The conditional distribution for the posterior function $\boldsymbol{f}_*$ can then be described as

$$\boldsymbol{f}_* | X, \boldsymbol{f}, X_* \sim \mathcal{N}\left(\overline{\boldsymbol{f}_*}, \operatorname{cov}(\boldsymbol{f}_*)\right),$$

where

$$\overline{\boldsymbol{f}_*} = \mathbb{E}[\boldsymbol{f}_* | X, \boldsymbol{f}, X_*] = K(X_*, X)[K(X, X)]^{-1}\boldsymbol{f}$$

$$\operatorname{cov}(\boldsymbol{f}_*) = K(X_*, X_*) - K(X_*, X)[K(X, X)]^{-1}K(X, X_*).$$

The function that we are looking for is $\overline{\boldsymbol{f}_*}$, whereas $\operatorname{cov}(\boldsymbol{f}_*)$ helps describe the uncertainty of the predictions. By choosing an apt covariance function for the prior distribution, it is simply a matter of training the model to find the hyperparameters that describe the covariance function of the posterior distribution. Then the mean function can be calculated.