Understanding of Gaussian Process

Problem statement:

Model a system with multiple variables using Gaussian process.

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

$$\Sigma = \mathrm{Cov}(X_i, X_j) = E\left[(X_i - \mu_i)(X_j - \mu_j)^T\right]$$

$$f \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \mu(x_1) \\ \mu(x_2) \\ \vdots \\ \mu(x_n) \end{bmatrix}}_{\text{Mean vector}}, \underbrace{\sigma^2 \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots \\ K(x_2, x_1) & K(x_2, x_2) & \\ \vdots & & \ddots \\ & & & K(x_n, x_n) \end{bmatrix}}_{\text{Covariance matrix}} \right)$$

To establish the Gaussian distribution of the data, the **mean vector** and **covariance matrix** are the two main outputs we are looking for.

Implementation - Theory

Introduction and calculation of kernel

The idea of the kernel is to combine the distribution of multiple variables in one system. The need for such a combination is that variables in one system influence each other, in other words, they are not independent.

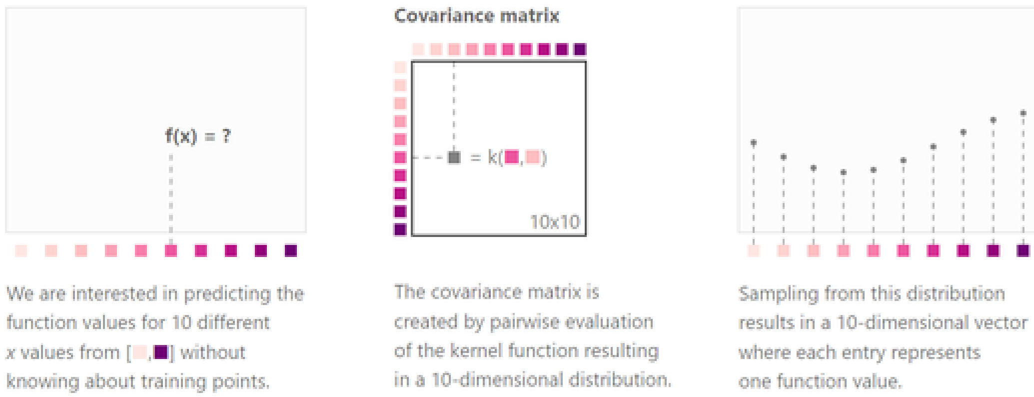
And the kernel itself is the energy function in GP waiting to be optimized, where its form of it is defined by both kernel type and hyperparameter.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{k=1}^{m} K(x_i^k, x_j^k)$$

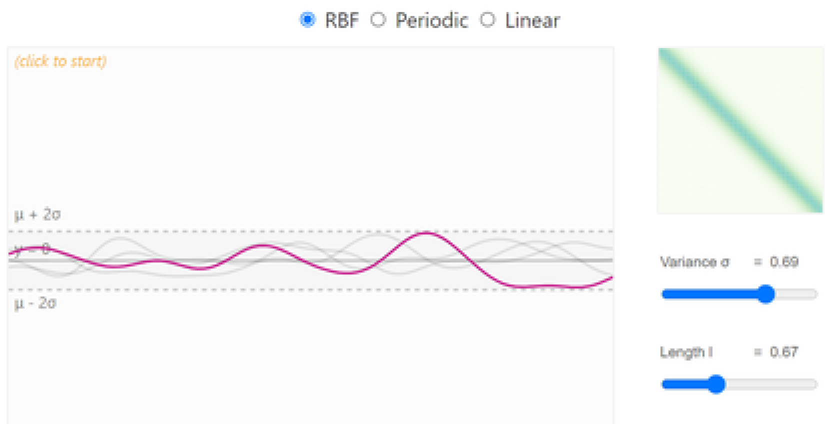
Here we use the most basic kernel:

RBF KERNEL

$$\sigma^2 \exp\left(-\frac{||t - t'||^2}{2l^2}\right)$$



Note the parameters - **Variance and Length** need to be decided and adjusted later.



Optimization of kernel

The process of optimization is to search the kernel that best fit the data(observation) , where the Bayes theory is in need.