Mathematics

Q.1 Gaussian Process Prior

↳ Lets assume the mean function to be 0. i.e ~~u(x)~~ $m(x)=0$ & the covariance function ~~k(x,x')~~

$$K(x,x') = \sigma^2 \times e^{\left(\frac{(x-x')^2}{2l^2}\right)}$$

$\sigma^2 \rightarrow$ Variance

$l \rightarrow$ length scale

This is the Radial Basis ~~Kernel~~ Function (RBF) kernel.

$$u \sim GP(0, K(x,x'))$$

$$p(u) = N(0,K)$$

Training data set is $S = \{(x_1, y_1) \ldots (x_n, y_n)\}$

$$y = u(x) + \varepsilon$$

~~$p(\varepsilon) = N(0, \sigma^2 I)$~~

Covariance Matrix

$$K(X,X) = \begin{bmatrix} k(x_1,x_1) & K(x_1,x_2) & \ldots & K(x_1,x_n) \\ K(x_2,x_1) & K(x_2,x_2) & \ldots & \\ \vdots & & & \\ K(x_n,x_1) & K(x_n,x_2) & \ldots & K(x_n,x_n) \end{bmatrix}$$

Now, as we want to interpolate the data points exactly. We will add some noise to the diagonal elements.

$$K(X,X) = K(X,X) + \sigma_n^2 I$$

Posterior Predictive Distribution

$$\cancel{\mu(x^*) = K(x^*,X) * x\, [K(X,X) + \sigma_n^2 I]}$$

$$\mu(x^*) = \frac{K(x^*,X) \times F(X)}{[K(X,X) + \sigma_n^2 I]}$$

$$\sigma^2(x^*) = K(x^*,x^*) - \mu(x^*)$$

Here, x^* is test data points.
$K(x^*,X)$ is a vector of covariances between the test data points & training data points.
$F(X)$ is the vector of function values corresponding ~~to~~ the training datapoints.

Using the posterior predictive distributions, we can interpolate between the given data points & extrapolate into the test data points. The mean function $\mu(x^*)$ provides the interpolated ~~data~~ values, while the variance function $\sigma^2(x^*)$ gives an indication of the uncertainity in the predictions.

Note that in our question there is no added noise so $\sigma^2_n I = 0$.

The ~~...~~ kernel function should be positive semidefinite so that we can perform cholesky decomposition on the matrix.

~~In Cholesky~~

To compute cholesky decomposition we need to find a lower triangular matrix L such that

$$K(X,X) = L \cdot L^T$$