Mathematics

Gaussian processes is an ~~infinitely~~ infinite dimentional multivariate gaussian ~~process~~ distribution whose subsets will be a multivariate gaussian distribution.

Given set of datapoints X & function $F(x)$.

~~Assuming~~ $y = F(x) \quad \forall\ x \in X$

$\vec{y} = [\, y(x_1), y(x_2) \; -- \; y(x_n) \,]$, X has n samples.

Assuming $\vec{y}$ ~~belongs to a gp~~ is a multivariate gaussian distribution which belongs to a gaussian process with zero mean ~~vector~~ function and a covariance function $K(x, x')$.

Here I have taken the covariance function the RBF ~~function~~ (Radial Basis Function) Kernel given as

$$K(x, x') = \sigma^2 \exp\left(-\frac{1}{2l^2} \|x - x'\|^2 \right), \quad x, x' \in X$$

σ: Output scale

l: length scale

As ~~Our~~ ~~cover~~ we know ~~the~~ the property of covariance matrix is to be positive semidefinite.

Thus a Kernal is vaild only when the Kernal matrix is positive-semidefinite for over a set of inputs $x_1, x_2 \ldots x_n \in X$.

So, assuming ~~a~~ ~~y~~ function y drawn from a zero mean Gaussian process with covariance function as RBF Kernal.

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

let y' be the predictive function Given x' as a unknown data for the same given distribution.

Using marginalization, data points x, x' ~~following~~ follows a multivariate joint gaussian distribution.

DATE
PAGE

$$\begin{bmatrix} y \\ y' \end{bmatrix} \sim N\left(0, \begin{bmatrix} K_{X,X} & K_{X,X'} \\ K_{X',X} & K_{X',X'} \end{bmatrix}\right)$$

partition matrix ~~between the covariance matrices~~

Using properties of Gaussian,

~~$\vec{y}_*$~~ $y' | y, X, X_*$ ~~with~~ will also be a gaussian distribution.

$$y' | y, X, X_* \sim N(\mu, \Sigma).$$

~~$\mu = K(X_*, K(X', X))(K6$~~

$$\mu = K_{X',X} \cdot K_{X,X} \cdot \vec{y}$$

$$\Sigma = K_{X',X'} - K_{X',X}(K_{X,X})^{-1} K_{X,X'}$$

The role of cholesky decomposition is to compute the inverse of the covariance matrix

to compute the ~~covaria~~ covariance matrix of the predictive posterior distribution.

Given covariance matrix Σ

$$\Sigma = L \cdot L^T$$

$L \rightarrow$ lower triangular matrix

$X \rightarrow$ Inverse of covariance matrix :-

$\Rightarrow \Sigma X = I$

$\Rightarrow L \cdot L^T X = I$

$(L^T X = B)$

$\Rightarrow LB = I \Rightarrow B = L^{-1}.$

$L^T X = B$

$\Rightarrow$

$L^T X = L^{-1}$

L^{-1} is also a lower triangular matrix with its diagonal element inverse of L.

for $i \leq j \leq N$. we do not compute B & instead use backward substitution.

and finding A^{-1}.