Research Module

Gewei Cao

December 14, 2022

1 Gaussain Process with Noise

1.1 Definition

We have data $\mathbf{D} = \{(\boldsymbol{x_i}, y_i)\}_{i=1}^{M}$, and assume that mean of y is 0.

Task: find the distribution of $f^*(x)$.

Assume that the true form of prediction function is: $y_i = f(\boldsymbol{x_i}) + \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. Here we have a M dimensional dependent variable $\mathbf{y}$, and a $M \times N$ dimensional independent variable $\mathbf{X}$, where M is the number of observations, and N is the dimension of x, i.e. $\boldsymbol{x_i} \sim \mathbb{R}^N$. The function $f(\boldsymbol{x_i}) : \mathbb{R}^N \to \mathbb{R}$ takes vector $\boldsymbol{x_i} \in \mathbb{R}^N$. Let $\boldsymbol{K}_{X,X} = k(\boldsymbol{x}, \boldsymbol{x}^T)$ which is the matrix of $k(\boldsymbol{x_i}, \boldsymbol{x_j})$. Thus, $\mathbf{K}$ is a $M \times M$ matrix.

The assumption of Gaussian Process is as following:
For a given vector $\mathbf{y}$, and its corresponding data $\mathbf{X}$, where vector $\boldsymbol{y} \in \mathbb{R}^M$ and $\boldsymbol{X}$ is $M \times N$ matrix. In addition, for $\mathbf{y}$ and $\mathbf{X}$ data, the error term $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma^\epsilon)$, and $\Sigma^\epsilon = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2, ..., \sigma_M^2)$. Meanwhile we have arbitrary $n \times N$ matrix $\mathbf{Z}$ and predicted value $f^*(\boldsymbol{z}) \in \mathbb{R}^n$, where $\boldsymbol{z} = (\boldsymbol{z_1, z_2, z_3, ..., z_n})^T$.

Then we assume $\mathbf{y}$ and $f^*(\boldsymbol{z})$ follow a $(M+n)$ multivariate normal distribution:

$$\begin{bmatrix} f^*(\boldsymbol{z}) \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{f^*(\boldsymbol{z})} \\ \mu_{\boldsymbol{y}} \end{bmatrix} , \begin{bmatrix} \boldsymbol{K}_{Z,Z} & \boldsymbol{K}_{Z,X} \\ \boldsymbol{K}_{X,Z} & \hat{\boldsymbol{K}}_{X,X} \end{bmatrix} \right) \tag{1}$$

where $\hat{\boldsymbol{K}}_{X,X} = \boldsymbol{K}_{X,X} + \Sigma^\epsilon$.

Then given data $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{Z}$, according to the conditional distributions of the multivariate normal distribution[1], we have the posterior distribution

$$f^*(\boldsymbol{z})|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z} \sim \mathcal{N}(\mu_{f^*(\boldsymbol{z})} + \boldsymbol{K}_{Z,X}\hat{\boldsymbol{K}}_{X,X}^{-1}(\boldsymbol{y} - \mu_{\boldsymbol{y}}), \boldsymbol{K}_{Z,Z} - \boldsymbol{K}_{Z,X}\hat{\boldsymbol{K}}_{X,X}^{-1}\boldsymbol{K}_{X,Z}) \tag{2}$$

[1] https://statproofbook.github.io/P/mvn-cond

1.2 Intuition behind Gaussain Process

The idea behind this process is that, assume our interested function is f(x) $f(\boldsymbol{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$, and we have an arbitorary vector of independent variable $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_M})^T$, and for each $\boldsymbol{x_i}, i = 1, 2, ..., M, x_i \in \mathbb{R}^N$, then we can obtain a series of $f(\boldsymbol{x}) = (f(\boldsymbol{x_1}), f(\boldsymbol{x_2}), f(\boldsymbol{x_3}), ..., f(\boldsymbol{x_M}))^T$. We assume that the series of f($\mathbf{x}$) follows a multivariate normal distribution which is:

$$f(\boldsymbol{x}) \sim \mathcal{N}(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}^T)) \tag{3}$$

This is the prior distribution of our function f(x), here we have a set of infinitely functions that follow this distribution, their mean is the function $\mu(\boldsymbol{x_i})$, and the variance of them is $k(\boldsymbol{x_i}, \boldsymbol{x_i}^T)$. Here we use kernel to denote variance-covariance matrix because kernel value represents how near two data points in the space are, with this property we can obtain a smooth function. Meanwhile, kernel matrix is symmetric and semi-positive-definite, this is needed for variance-covariance matrix.

When we provide a set of observed data $D = \{\boldsymbol{x_i}, y_i\}_{i=1}^M$, and introduce noise term, then we have $y_i = f(\boldsymbol{x_i}) + \epsilon_i$, the property of noise term ϵ is the same as abouve. Data D is regarded as training data, and if we now have a testing data $\boldsymbol{x^*}$, the joint distribution of $(y_i, f(\boldsymbol{x^*}))$ is also multi-normal, here we can obtain the marginal distribution of $f(\boldsymbol{x^*})$, which is the conditional distribution given data D, and this is called posterior distribution. Thus, we also obtain the posterior mean function and posterior variance function of $f(\boldsymbol{x^*})$.

If our test data set contains multiple data, then we will obtain a vector of $f^*(\boldsymbol{x^*})$ and its posterior multivariate normal distribution as mentioned in section 1.1.

1.3 Gaussain Process in research paper

In our object paper, for given price data P,