统计学习第七章

贾金柱

Dec 7, 2017

讲课老师

- 主讲人： 贾金柱
  - Email: jzjia@math.pku.edu.cn
- 助教：肖一君
  - Email: xiaoyijun1994@126.com

第七章 Gaussian Process and its applications

- 课程目标
    - Gaussian Process methods for Regression

Gaussian Process

定义 [GP] A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

- 多元正态分布有均值和协方差矩阵刻画
- GP 由 mean function $m(x)$ 和 covariance function $k(x, x')$刻画

$$m(x) = E[f(x)]$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$$

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')).$$

Bayesian linear model

- Bayesian linear model is a GP
- 考虑线性模型 $f(x) = \phi(x)^T w$, 假设先验 $w \sim N(0, \Sigma_p)$.
- $f(x)$ 就是一个 GP

$$m(x) = E(f(x)) = \phi(x)^T E(w) = 0$$

$$k(x, x') = cov(f(x), f(x')) = \phi(x)^T \Sigma_p \phi(x')$$

更一般的GP

- Bayesian linear model 的缺点是：需要事先制定一个基底 $\phi(x)$
- 可以通过指定 $k(x, x')$来避开 基底$\phi(x)$的指定。
- Suppose $f(x) \sim GP(0, k(x, x'))$
- 常用的kernel 有Gaussian Kernel $k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\lambda}\right)$

使用 GP 做预测

- 先看 Noise-free 情形

Given train data $(x_i, f_i), i = 1, 2, \dots, n$. 求 test data x^* 处的 f 的值 $f(x^*)$ 是多少?

注意到

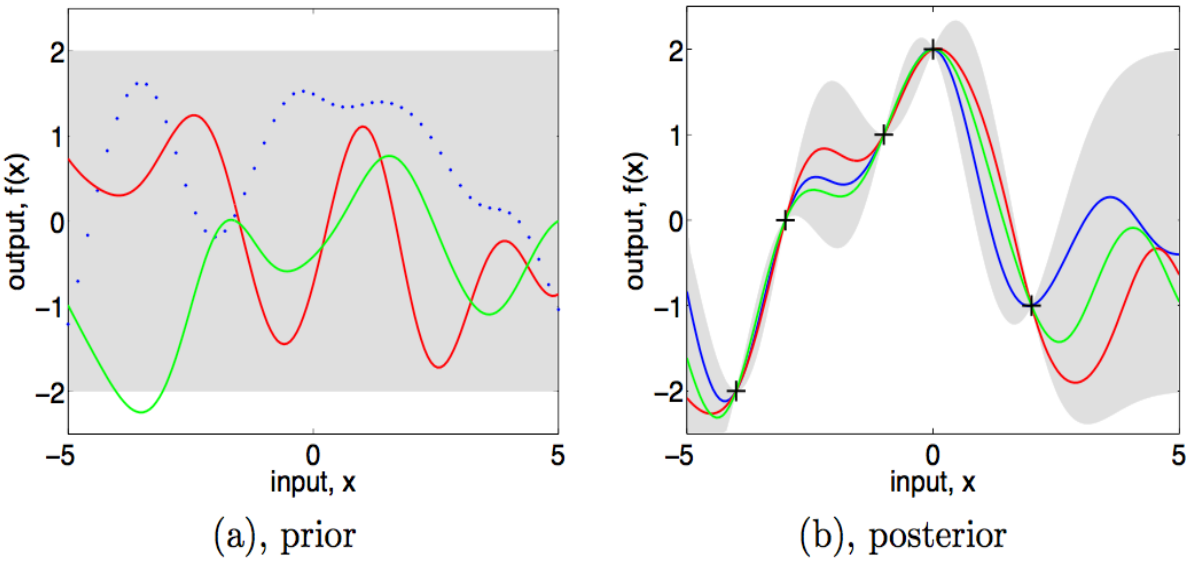
$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim N\Big(0, \begin{bmatrix} K(X,X), K(X,X^*) \\ K(X^*,X), K(X^*,X^*) \end{bmatrix}\Big)$$

所以,

$$f^*|X^*, X, f \sim N\Big(K(X^*,X)K(X,X)^{-1}f,$$
$$K(X^*,X^*) - K(X^*,X)K(X,X)^{-1}K(X,X^*))$$

使用 GP 做预测 (one example)

- prior and posterior of GP



(a), prior (b), posterior

图: prior and posterior of GP

使用 GP 做预测

- 再看有noise 的情形
- $y_i = f(x_i) + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$.

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim N\Big(0, \begin{bmatrix} K(X,X) + \sigma^2 I, K(X,X^*) \\ K(X^*,X), K(X^*,X^*) \end{bmatrix}\Big)$$

所以,

$$f^*|X^*, X, f \sim N\Big(K(X^*,X)[K(X,X) + \sigma^2 I]^{-1} f,$$
$$K(X^*,X^*) - K(X^*,X)[K(X,X) + \sigma^2 I]^{-1} K(X,X^*))$$