假设有3个点 x_1 x_2 x_3 $\longrightarrow f(x_1), f(x_2), f(x_3)$
我们通过GPR(高斯过程回归)得到 x_4 的值.

问题描述

高斯回归关键假设:

1. 对于给定的 x, 得到 y 这些 y 服从联合正态分布.

$$f(x_1) = f_1 = y_1$$
$$f(x_2) = f_2 = y_2$$
$$f(x_3) = f_3 = y_3$$

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}\right)$$

2. 如果 x 之间的距离, y 之间的相似度高
这些 y 的相似度用 k 的协方差表示.

如何衡量? 满足如上假设
(半正定)

SVM中的核, 即满足 Mecer kernel

比如常见高斯核 $k = \exp(-\lambda \|x_i - x_j\|^2)$

对已知数据的高斯假定过程已完成，
新数据如何计算？x_4？

$$\begin{bmatrix} \vec{f} \\ f_4 \end{bmatrix} \sim \left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix} \right]$$

k $\qquad$ k_4

$$\begin{cases} \vec{f} \sim N(0, k) \\ [f_1, f_2, f_3]^T = \vec{f} \end{cases}$$

k是可以直接算出来的
k_4部分也通过核方法算出来的

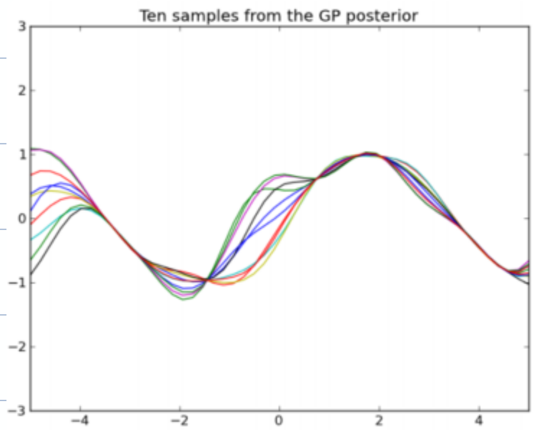
所以(f, f_4)的联合分布$P(f, f_4)$所有参数已知，根据边际分布+条件分布可得：

$$f_4 \sim N(\mu_4, \sigma_4)$$

where $\begin{cases} \mu_4 = k_4^T k^{-1} \vec{f} \\ \sigma_4 = -k_4^T k^{-1} k_4 + k_{44} \end{cases}$

※ f_4为分布，不是点估计.

为了得到真实的每个新样本的 output，采取采样的形式



采样10次

问题1、如何到样本 output？

直接把若干次采样的结果平均
即可作为一次拟合曲线.

问题2：采样图有什么特点？

有已知点（x_1, x_2, x_3）地方方差小
在不确定预测点的时候关注方差+期望
（参考 UCB 理论）.

以上为过程探索

总结：

$$f(x) \sim GP(\mu(x), k(x, x'))$$

where $\begin{cases} \mu(x) = E(f(x)) \\ k(x, x') = E[(f(x)-\mu(x))(f(x')-\mu(x'))] \end{cases}$

核函数

有训练集 $D\ \{x_i, f(x_i)\} = \{x_i, f_i\}$

求测试集 X' ($N' \times D$)

D 维度.

$$\begin{bmatrix} \vec{f} \\ \vec{f'} \end{bmatrix} \sim N\left(\begin{pmatrix} \mu \\ \mu' \end{pmatrix}, \begin{bmatrix} k, & k' \\ k', & k'' \end{bmatrix} \right)$$

$$k_{ij} = k(i, j) = \sigma'^2 \exp\left(-\frac{1}{2\sigma^2}(i-j)^2\right)$$

$$p(f' \mid x', X, f) = N(\mu', \Sigma')$$

where $\mu' = \mu(X') + k'^T k^{-1}(f - \mu(X))$

$\Sigma' = k'' - k'^T k^{-1} k'$