

周报

刘精昌

2016 年 10 月 7 日

本周工作

1. 阅读《Automatic Generation of Probabilistic Programming from Time Series Data》，基于 Compositional Kernel 工作，用概率编程语言 *Stan* 对相关工作进行描述。
2. 阅读《Gaussian Process Kernels for Pattern Discovery and Extrapolation》ICML13，论文引入一种 gp kernel——spectral mixture (SM) kernel，好处是对一些问题精度较高，最主要的是能 Pattern Discovery。

在信号分析等很多问题中，会对原始数据进行傅里叶变换，将时域上的数据转换到频域，通过分析频域上的高峰以得到相关信息。既然关注的是频域的高峰，而 gauss mixture 分布恰是多峰分布，可以先定义频域上的分布（谱密度）是 gauss mixture 分布。再经由傅里叶反变换，得到

$$k(\tau) = \sum_{q=1}^Q w_q \prod_{p=1}^P \exp\left(-2\pi^2 \tau_p^2 v_q^{(p)}\right) \cos\left(2\pi \tau_p \mu_q^{(p)}\right)$$

$\tau = x - x'$ ，便得到了 SM kernel。

如图 1 所示，左边是用不同的 kernel 对一个时间序列进行预测，蓝线是训练集，绿线是实际值，黑线是 SM 核的 GP 预测结果，其他线是用其他 kernel 的预测结果。右边是对应的对数谱密度，开始时定义了 10 个组件，训练后保留了 7 个组件。对数谱密度第一个 peak 在 0.00148 处，对应于时间序列的上升趋势。对于 0.08 处的 peak， $1/0.08=12$ ，刚好对应左边时间序列的一年周期。同样，其他的 peak 对应与左边时间序列的周期性趋势有关，具体数据可以做具体解释。peak 越尖锐，对应 gauss 分布方差越小，则周期趋势越显著。

下周计划

- 继续看 LDA 相关知识，看 MLAPP，处理课后习题，看看优化相关的论文。

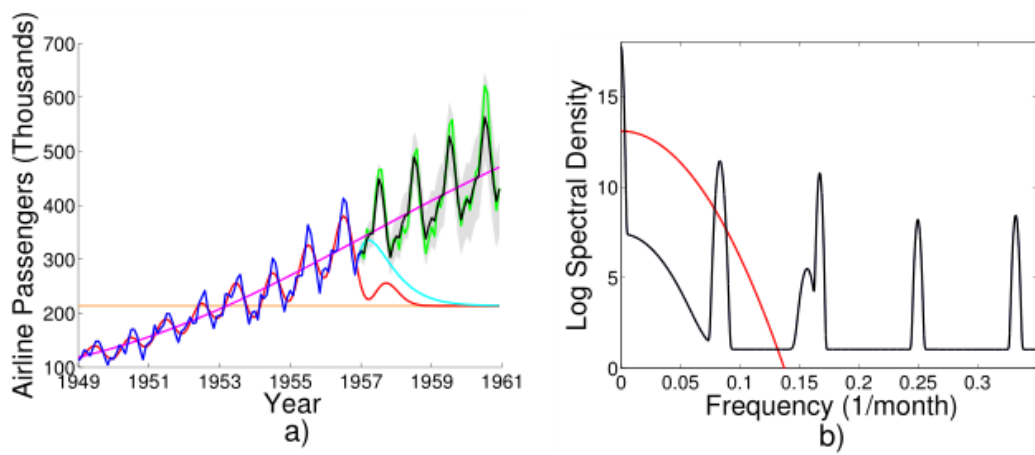


图 1: GP 预测及谱密度示意图