

【题目】高斯过程回归与股票预测

大家对于回归并不陌生。普通的线性回归采用如下模型：

$$Y = AX + N$$

其中 Y 是观测变量， X 为待求变量， N 为加性噪声， A 为已知的模型矩阵。

标准的线性最小二乘方法可以有效地解决该问题。如果将模型扩展到非线性，

$$Y = f(X) + N$$

可以采用非线性最小二乘方法进行求解。应该指出，无论是线性模型还是非线性模型，本质上的差异很小，原因在于模型本身（矩阵 A 或者函数 $f(x)$ ）是确定的对象，没有随机性。

近年来,人工智能和机器学习领域的学者将更多的目光投向了所谓“随机回归模型”，高斯过程回归（ Gaussian Processes Regression， GPR）就是其中的典型代表。 GPR 采用如下模型

$$Y = g(X) + N$$

其中 $g(X)$ 是一个和 X 存在紧密联系的高斯过程的样本轨道片段。准确地说， $g(X)$ 是和 X 有关的某个多元高斯分布的采样值。这样一来，回归的模型中引入了随机性（而不仅仅限于加性的观测噪声），回归也发生了本质的变化。

[1]是有关 GPR 最早的文献， [2]对 GPR 进行了十分简洁明快的介绍。 [3]和[4]是两份陈述直观的课件。 [5]和[6]是国外大学生做的 Project 报告，将 GPR 用于股票价格预测。 [7]和[8]给出了 GPR 在不同领域的应用实例，进一步说明其广泛的适用性。 [9]体现了 GPR 在短期时间序列预测方面的研究成果。 [10]是 Stanford 大学关于 Gaussian 过程的学习材料，其中 GPR 方面的介绍很清晰。 [11] 是卡内基梅隆大学的一篇博士论文，将 GPR 延伸到了非平稳高斯过

程上。

我们的任务是，在学习和消化 GPR 的基本知识的基础上，研究其在股票预测方面的效能，探索其在时间序列预测方面的应用潜力。

我们的数据来源是：

<http://www.nasdaq.com/>

鼓励同学们采用国外学生们研究过的几种股票，和他们比试一下，谁的预测模型更好，更加精确且效率更高。

关于 GPR 更多的材料，可以从下述网址得到：

<http://www.gaussianprocess.org/>

【参考文献】

- [1] C.K. I. Williams, C.E. Rasmussen "Gaussian Processes for Regression"
- [2] M.Ebden, " Gaussian Processes for Regression An Quick Introduction".
- [3] H.M.Wallace, "Introduction to Gaussian Process Regression"
- [4] Z. Ghahramani, " A Tutorial on Gaussian Processes"
- [5] M.T. Farrell, et al, "Gaussian Process Regression Models for Predicting Stock Trends".
- [6] Long-term Stock Market Forecasting using Gaussian Processes
- [7] J.M, Tomczak, et, al, "Gaussian process regression as a predictive model for Quality-of-Service in Web service systems"
- [8] Y. Altmann, et,al, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes".
- [9] H. Topa, et,al, "Gaussian process modelling of multiple short time series".
- [10] Chuong B. Do, "Gaussian Processes"
- [11] C.J. Paciorek, "NONSTATIONARY GAUSSIAN PROCESSES FOR REGRESSION AND SPATIAL MODELING"