Gaussian process regression methods and extensions for stock market prediction

Zexun Chen

Department of Mathematics
University of Leicester

zc74@le.ac.uk

PhD Viva

June 22, 2017

Thanks to you, my kind examiners:

Prof. Sergei Fedotov

Prof. Jeremy Levesley

Thanks to my kind supervisors:

Prof. Alexander N. Gorban

Dr. Bo Wang

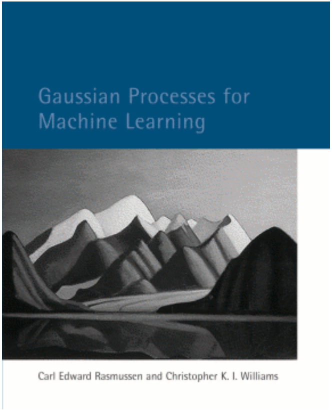
Overview

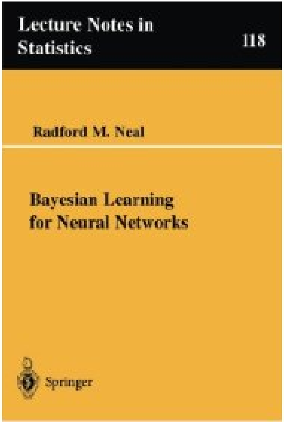
1. Thesis structure
2. Gaussian process regression (GPR)
   - Kernel and mean function
   - Parameter estimation
3. Gaussian process regression extensions
   - GPR extensions
   - Applications of GPR and its extensions
4. Further extensions for multi-output prediction
   - Multivariate Gaussian process regression (MV-GPR)
   - Multivariate Student$-t$ process regression (MV-TPR)
   - Applications of MV-GPR and MV-TPR
5. Summary

Thesis structure

- Gaussian process regression (GPR)
    - Model derivations, covariance function (kernel) and mean function
    - **Parameter estimation**
        - **Initial hyper-parameter selection (my work)**
- Gaussian process regression extensions and applications
    - Model extensions, including Gaussian process regression with Student$-t$ likelihood (GPRT), Student$-t$ process regression (TPR), state-space GPR (SSGPR), state-space TPR (SSTPR)
    - **Equity index prediction (my work)**
- **GPR extensions for multi-output prediction (my work)**
    - Multivariate Gaussian process regression (MV-GPR)
    - Multivariate Student$-t$ process regression (MV-TPR)
    - Simulated examples and stock market investment

Gaussian process for machine learning





Gaussian process

A Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distribution.

Theorem (Gaussian process)

For any set S, any mean function $\mu : S \mapsto \mathbb{R}$ and any covariance function $k : S \times S \mapsto \mathbb{R}$, there exists a GP $f(t)$ on S, s.t., $\mathbb{E}[f(t)] = \mu(t)$, $\mathrm{cov}(f(s), f(t)) = k(s, t), \forall s, t \in S$. It denotes $f(t) \sim \mathcal{GP}(\mu, k)$.

A collection of functions $[f(t_1), f(t_2), \ldots, f(t_n)]$ are realised through a multivariate Gaussian distribution

$$[f(t_1), f(t_2), \ldots, f(t_n)]^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, K)$$

where $\mu_i = \mu(t_i)$ and $K_{ij} = k(t_i, t_j)$.

Gaussian process regression

- Consider a noisy model $y = f(x) + \varepsilon$, where f is drawn from a GP and ε is Gaussian noise. Then, we wish to predict new observations $\boldsymbol{y}_*$ for new test points Z.
- The joint distribution of training observations $\boldsymbol{y}$ and test data $\boldsymbol{y}_*$ is

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(X) \\ \mu(Z) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(Z,X)^{\mathrm{T}} \\ K(Z,X) & K(Z,Z) \end{bmatrix} \right),$$

- The predictive distribution is $p(\boldsymbol{y}_* | X, \boldsymbol{y}, Z) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where

$$\hat{\boldsymbol{\mu}} = K(Z,X)^{\mathrm{T}}(K(X,X) + \sigma_n^2 I)^{-1}(\boldsymbol{y} - \mu(X)) + \mu(Z), \tag{1}$$

$$\hat{\boldsymbol{\Sigma}} = K(Z,Z) - K(Z,X)^{\mathrm{T}}(K(X,X) + \sigma_n^2 I)^{-1} K(Z,X). \tag{2}$$

Kernel and mean function

Kernels

- Basic kernels, such as, linear (LIN), squared exponential (SE), periodic (PER), etc.
- Composite kernels, such as local periodic (LP), spectral mixture (SM), etc. [Wilson, 2014]

Mean functions

- Zero-offset (mainstream) [Roberts et al., 2013]
- Linear
- Others

Maximum marginal likelihood

Sensitivity of initial hyper-parameters

- Negative log marginal likelihood (nlml) is not always convex, it may suffer a local optima
- Initial hyper-parameters selection may affect the final result

A common approach to overcome the sensitivity

- Randomly select several initial hyper-parameters in a reasonable range
- Compute nlmls for these initial hyper-parameters
- Select the best initial hyper-parameters with the smallest nlml

Problem: how to randomly select ?

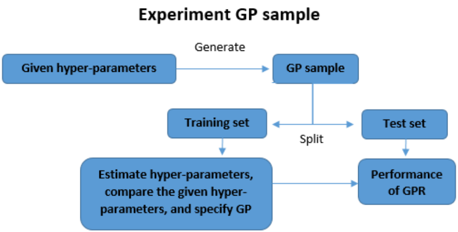
Problem: how to randomly select ?

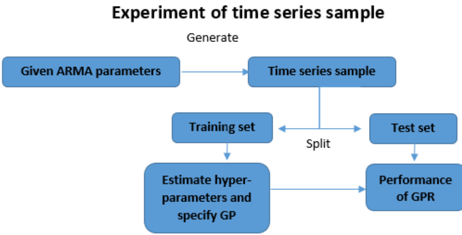
- Researchers' expert opinions and experiences
- Uniform (0,1) (mainstream)
- $\mathcal{N}(0,1)$
- Other simple distributions

My work

- **Put different priors on initial hyper-parameters**
- **Explore how priors of initial parameters affect the estimation**
- **Study how priors of initial parameters affect the performance of GPR model**

How priors of initial hyper-parameters affect GPR models





Experimental results [Chen and Wang, 2016]

- The hyper-parameter estimation depends on the choice of kernels
    - SE: robust
    - PER: very sensitive
- The performance of GPR is considerable regardless of the prior choice

GPR extensions

Fat-tailed distributions

- Student$-t$ distribution
- Pareto distribution
- Lévy distribution (Lévy walk [Fedotov and Korabel, 2017])
- Other stable distributions

Gaussian process regression with Student$-t$ likelihood (GPRT)

$y = f + \varepsilon$, where f follows a GP and ε follows Student$-t$ distribution [Vanhatalo et al., 2009].

Student$-t$ process regression (TPR)

$y = f$, where f follows a Student$-t$ process (TP) [Shah et al., 2014].

State-space GPR (SSGPR) and state-space TPR (SSTPR)

Consider the historical prices/indices as state in GPR and TPR model.

Some applications (my work)

Equity index series predictions with fixed forecasting horizon

- INDU, NDX, SPX, and UKX predictions using GPR, GPRT, and TPR
- DAX, HSI, INDU, NDX, NKY, SENSEX, SPX, UKX predictions using GPR, GPRT, TPR, and ARMA model

Result: GPR and TPR perform relatively better overall

Model validation based on equity index predictions

- Objective: ten equity indices, including DAX, HSI, INDU, NDX, NKY, SENSEX, SHSZ300, SPX, UKX, XU100.
- Methods: leave-one-out cross-validation, k-fold validation ($k = 10$) and sliding windows (using GPR and TPR).
- **Result: Overall both GPR and TPR perform well and the superiority of them depends on the data set.**

Some applications (my work)

State-space models for equity index predictions

- Objective: ten equity indices, including DAX, HSI, INDU, NDX, NKY, SENSEX, SHSZ300, SPX, UKX, XU100.
- Methods: sliding windows using State-space Gaussian process regression (SSGPR) and State-space Student$-t$ process regression (SSTPR) [Brahim-Belhouari and Bermak, 2004, Girard et al., 2003]
- **Result: SSTPR performs better than SSGPR overall**

Stock market efficiency analysis

- Data: sliding windows predictions using SSGPR and SSTPR
- **Result: the developed markets are efficient on the whole while sometimes some emerging markets are also unpredictable, for example, Chinese market**

Multi-output prediction

- Chinese company in NASDAQ
- Internet sector

How can we make multi-output predictions by extending GPR?

Gaussian process for vector-valued function

Main steps in the existing methods [Boyle and Frean, 2005, Alvarez et al., 2011, Wang and Chen, 2015]

- Vectorise the multi-output variables(convert a matrix to a vector) $\mathrm{vec}[(\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_n))^{\mathrm{T}}]$, where $\boldsymbol{f} : \mathcal{X} \mapsto \mathbb{R}^d$ is a vector-valued function and $\mathrm{vec}[\cdot]$ is vector operator
- Construct a new "big" covariance matrix, usually Kronecker product, $\Omega \otimes K$), where $K \in \mathbb{R}^{n \times n}$ is still a covariance matrix and $\Omega \in \mathbb{R}^{d \times d}$ captures the covariance of outputs
- Assume $\mathrm{vec}[(\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_n))^{\mathrm{T}}] \sim \mathcal{N}(\mathbf{0}, \Omega \otimes K)$
- Operate the remaining steps similar to the classical GPR

Essential assumption

$$(\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{x}_n))^{\mathrm{T}} \sim \mathcal{MN}(\mathbf{0}, K, \Omega).$$

Problem and solution

Essential property

Matrix-variate Gaussian distribution can be reformulated as multivariate Gaussian distribution [Gupta and Nagar, 1999].

Can we use this method to extend multi-output TPR? **NO !**
Can we have another way to derive multi-output GPR, which can be easily extended to multi-output TPR? **YES !**

Standard solution of GPR and TPR for multi-output prediction (my work)

- Define a multivariate Gaussian (Student$-t$) process
- Carry out regression model (matrix form) with the assumption of the multivariate Gaussian (Student$-t$) process directly
- Estimate the parameters directly (matrix computation)
- Make predictions directly (matrix computation)

Multivariate Gaussian process regression (MV-GPR)

Given n pairs of observations $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n, \boldsymbol{x}_i \in \mathbb{R}^p, \boldsymbol{y}_i \in \mathbb{R}^d$, assuming

$$\begin{aligned} \boldsymbol{f} &\sim \mathcal{MGP}(\mathbf{0}, k', \Omega), \\ \boldsymbol{y}_i &= \boldsymbol{f}(\boldsymbol{x}_i), \text{for } i = 1, \cdots, n, \end{aligned}$$

where $k' = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \delta_{ij}\sigma_n^2$, $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$. The joint distribution of the training observations Y and the predictive targets $\boldsymbol{y}_*$

$$\begin{bmatrix} Y \\ \boldsymbol{y}_* \end{bmatrix} \sim \mathcal{MN}\left(\mathbf{0}, \begin{bmatrix} K'(X,X) & K'(Z,X)^{\mathrm{T}} \\ K'(Z,X) & K'(Z,Z) \end{bmatrix}, \Omega\right), \tag{3}$$

Therefore, the predictive distribution is

$$p(\boldsymbol{y}_*|X, Y, Z) = \mathcal{MN}(\hat{M}, \hat{\Sigma}, \hat{\Omega}), \tag{4}$$

where

$$\begin{aligned} \hat{M} &= K'(Z,X)^{\mathrm{T}} K'(X,X)^{-1} Y, & (5) \\ \hat{\Sigma} &= K'(Z,Z) - K'(Z,X)^{\mathrm{T}} K'(X,X)^{-1} K'(Z,X), & (6) \\ \hat{\Omega} &= \Omega. & (7) \end{aligned}$$

Comparison

Table 1: The comparison of existing multi-output GPR and MV-GPR

	Existing multi-output GPR	MV-GPR
Predictive distribution	Same	
Formulation	Complicated	Concise
Extension	Non-exist	MV-TPR

Multivariate Student−t process regression (MV-TPR)

Given n pairs of observations $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}, \boldsymbol{x}_i \in \mathbb{R}^p, \boldsymbol{y}_i \in \mathbb{R}^d$, we assume

$$\begin{aligned} \boldsymbol{f} &\sim \mathcal{MTP}(\nu, \boldsymbol{u}, k', \Omega), \nu > 2, \\ \boldsymbol{y}_i &= \boldsymbol{f}(\boldsymbol{x}_i), \text{for } i = 1, \cdots, n, \end{aligned}$$

Therefore the predictive distribution is

$$p(\boldsymbol{f}_*|X, Y, Z) = \mathcal{MT}(\hat{\nu}, \hat{M}, \hat{\Sigma}, \hat{\Omega}), \tag{8}$$

where

$$\begin{aligned} \hat{\nu} &= \nu + n, && (9) \\ \hat{M} &= K'(Z, X)^{\mathrm{T}} K'(X, X)^{-1} \boldsymbol{y}, && (10) \\ \hat{\Sigma} &= K'(Z, Z) - K'(Z, X)^{\mathrm{T}} K'(X, X)^{-1} K'(Z, X), && (11) \\ \hat{\Omega} &= \Omega + Y^{\mathrm{T}} K'(X, X)^{-1} Y. && (12) \end{aligned}$$

Experiments and results

- Simulated data prediction

	Simulated data with correlated	
	Gaussian noise	Student$-t$ noise
Perform	MV-TPR>MV-GPR>TPR>GPR	MV-TPR>MV-GPR>TPR>GPR

- Stock market investments

(1)	Three Chinese stocks investment in NASDAQ		
	BIDU	NTES	CTRP
Result	Insignificant difference	Insignificant difference	MV-TPR best

(2)	Stocks investment from diverse sectors in Dow 30						
	Oil&Gas	Industrial	Goods	Health Care	Services	Financials	Technology
Best model	TPR	TPR	MV-GPR	MV-TPR	MV-GPR	MV-GPR	GPR

What have I done in this thesis?

- Reviews
  - GPR (model derivation, kernel, mean function, parameter estimation)
  - GPR extensions, including TPR, GPRT, SSGPR and SSTPR
- Further studies and developments
  - Initial hyper-parameters selection (put different priors on the initials)
  - Another derivation of the existing dependent Gaussian process (MV-GPR)
  - New considerable method for multi-output prediction (MV-TPR)
- Applications
  - Predict many equity indices (GPR and its extensions)
  - Make stock market investment based on Buy&Sell strategy (MV-GPR and MV-TPR)

What have I achieved in this thesis?

- Study the sensitivity of initial hyper-parameters and find that
  - the initial hyper-parameters' estimates depend on the choice of kernel
  - the performance of GPR is satisfactory in terms of predictability regardless of the prior selection
- Apply GPR and its extension to equity index prediction and find that
  - GPR and TPR have relatively considerable capability of predicting performances
  - the developed markets are efficient overall
- Propose MV-GPR and MV-TPR and find that both MV-GPR and MV-TPR
  - have considerable predictions on the simulated data, particularly MV-TPR performs better
  - are shown profitable in stock market investment

References I

Alvarez, M. A., Rosasco, L., and Lawrence, N. D. (2011).
Kernels for vector-valued functions: A review.
arXiv preprint arXiv:1106.6251.

Boyle, P. and Frean, M. (2005).
Multiple output Gaussian process regression.

Brahim-Belhouari, S. and Bermak, A. (2004).
Gaussian process for nonstationary time series prediction.
Computational Statistics & Data Analysis, 47(4):705–712.

Chen, Z. and Wang, B. (2016).
How priors of initial hyperparameters affect gaussian process regression models.
arXiv preprint arXiv:1605.07906.

Chen, Z., Wang, B., and Gorban, A. N. (2017).
Multivariate gaussian and student $-t$ process regression for multi-output prediction.
arXiv preprint arXiv:1703.04455.

Fedotov, S. and Korabel, N. (2017).
Emergence of Lévy walks in systems of interacting individuals.
Physical Review E, 95(3):030107.

References II

Girard, A., Rasmussen, C., Quiñonero Candela, J., and Murray Smith, R. (2003).
Gaussian process priors with uncertain inputs-application to multiple-step ahead time series forecasting.
Advances in Neural Information Processing Systems, pages 545–552.

Gupta, A. K. and Nagar, D. K. (1999).
Matrix variate distributions, volume 104.
CRC Press.

MacKay, D. J. (1998).
Introduction to Gaussian processes.
NATO ASI Series F Computer and Systems Sciences, 168:133–166.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., and Aigrain, S. (2013).
Gaussian processes for time-series modelling.
Phil. Trans. R. Soc. A, 371(1984):20110550.

Shah, A., Wilson, A. G., and Ghahramani, Z. (2014).
Student-t processes as alternatives to Gaussian processes.
arXiv preprint arXiv:1402.4306.

References III

Vanhatalo, J., Jylänki, P., and Vehtari, A. (2009).
Gaussian process regression with Student-t likelihood.
In *Advances in Neural Information Processing Systems*, pages 1910–1918.

Wang, B. and Chen, T. (2015).
Gaussian process regression with multiple response variables.
Chemometrics and Intelligent Laboratory Systems, 142:159–165.

Wilson, A. G. (2014).
Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes.
PhD thesis, PhD thesis, University of Cambridge.


THANK YOU
for your
ATTENTION!