Gaussian Process and its Application in Engineering

Hao Chen [hao.chen@stat.ubc.ca]
Department of Statistics, University of British Columbia [www.stat.ubc.ca]

The First Overseas Young Scholars Forum
Nanjing University of Posts and Telecommunications [www.njupt.edu.cn]
Outline

1. On-going Research (big picture)
   - Gaussian process (GP).
   - Sequential experimental design (active learning algorithm).

2. Proposed Future Research (big picture)
   - Gaussian process classification (GPC)
   - Statistical modelling and machine learning.

3. Potential Contributions

4. Comments and Feedback
Some Basics

- **Model v.s. Method**

  \[ y = \beta_0 + \beta_1 x \]

  (1) Maximum Likelihood Estimation
  (2) Least Square Estimation
  (3) Method of Moment

Method: The way to guess a number s.t. it has “good” properties:
(1) Unbiased: \( E(\hat{\beta}_0) = \beta_0 \)
(2) Consistent: \( \lim_{n \to \infty} P(|\hat{\beta}_0 - \beta_0| > \epsilon) = 0. \)

Parameter Estimation: Repeated Experiments and Large Sample Size.
Some Basics

The difference between supervised machine learning and unsupervised machine learning.

- **Supervised** learning: Learn a model based on the available data and make **predictions** of $y$ for new future untried data:

  - **Classification**: $y$ is categorical: handwriting recognition of numbers 0-9 (10 categories); financial equity direction, up or down (2 categories), .......
  - **Regression**: $y$ is numerical: weather forecast (tomorrow's precipitation, time series analysis), housing price forecast....
The difference between supervised machine learning and unsupervised machine learning.

- **Supervised** learning: Learn a model based on the available data and make **predictions** of $y$ for new future untried data:
  - **Classification**: $y$ is categorical: handwriting recognition of numbers 0 – 9 (10 categories); financial equity direction, up or down (2 categories), .......

- **Unsupervised** learning: Learn the structure/pattern of the data itself to have a better understanding. NOT to make predictions.
  - Cluster Analysis is a typical unsupervised learning.
Some Basics

The difference between supervised machine learning and unsupervised machine learning.

- **Supervised** learning: Learn a model based on the available data and make **predictions** of $y$ for new future untried data:
  - **Classification**: $y$ is categorical: handwriting recognition of numbers 0 – 9 (10 categories); financial equity direction, up or down (2 categories), .......
  - **Regression**: $y$ is numerical: weather forecast (tomorrow’s precipitation, time series analysis), housing price forecast....
Some Basics

The difference between supervised machine learning and unsupervised machine learning.

- **Supervised** learning: Learn a model based on the available data and make *predictions* of $y$ for new future untried data:
  - **Classification**: $y$ is categorical: handwriting recognition of numbers 0 – 9 (10 categories); financial equity direction, up or down (2 categories), ........
  - **Regression**: $y$ is numerical: weather forecast (tomorrow’s precipitation, time series analysis), housing price forecast....

- **Unsupervised** learning: Learn the *structure/pattern* of the data itself to have a better understanding. NOT to make predictions. **Cluster Analysis** is a typical unsupervised learning.
Gaussian Process. A stochastic regression model, i.e., $y$ is numerical.

- The effect of regression term and correlation structure on its prediction accuracy. *Statistical Science*
- The effect of design on its prediction accuracy. Ongoing
- Bayesian parameter estimation of a Gaussian process based on Metropolis-Hastings algorithm (MH is a typical Markov chain Monte Carlo, (MCMC)) *SIAM/ASA Journal on Uncertainty Quantification*
On-going Research

Sequential experimental design (active learning algorithm). \( n = 30 \)

1. Step 1: Use part of the data \( (n_0 = 20) \) to get an initial estimate.
2. Step 2: Add a point to the training set by optimizing a search criterion.
3. Step 3: Once add a point, update the estimate.
4. Iterate between Step 1 and Step 3 until the budget has been exhausted.

- Constrained Optimization. *Technometrics*
- Rare Failure Probability Estimation. *Ongoing, will submit to Journal of Quality Technology*
On-going Research

(a) Initial design: random

(b) Initial design: random

Figure:
1. GP classification: parameter estimation; assumption violation, .......
3. Statistical modelling and data analysis: collaboration with other department (Dr. Zhou Liang’s presentation)
Potential Contributions

1. Research: Publications on top-notch journals.
2. Teaching: (1) Definitions, concepts and foundation. (2) Systematical study.
THANK YOU! 😊