**National University of Singapore**
**School of Computing**
**CS4246 AI Planning and Decision Making**

# Term Project: Planning and Decision Making in the Real World
# Part 1: Gaussian Process (GP) Modeling

Issue: September 1, 2016 Due: October 6, 2016

## Important Instructions

1. *Your project report must be TYPE-WRITTEN using the latest AAAI Press Word template or LaTeX macro located in the author kit*[1]*. It should NOT be more than* 6 *pages (inclusive of diagrams, references, and team details described in bullet point* 5*).*
2. *Submit your project report in PDF format to the workbin folder named "Project Report 1 Submission" in IVLE by* **October 6, 2016, 11:59pm***. Late submissions will NOT be accepted.*
3. *You should submit ONE project report per team.*
4. *Each team should have 5 to 6 members.*
5. *Indicate clearly your team number and all names and matriculation numbers of the members of the team on the first page of the project report.*
6. *In a separate section, describe clearly the role(s) of each team member in this project.*
7. *Note that we may call any one member of a team to explain the submitted report during the marking process. You are advised NOT to free-ride.*

All members of the team will be given the same grade/mark[2]. The grading criteria is as follows:

[1] http://www.aaai.org/Publications/Templates/AuthorKit.zip
[2] Free riders are given zero.

| | |
|---|---|
| 5% | **Motivating application**: Describe clearly and CONCISELY a MOTIVATING REAL-WORLD application scenario detailing why there is a need for using the rich class of *Gaussian process* (GP) models for probabilistic non-linear regression and prediction. Doing so will allow you to demonstrate the practical significance of your research project work. For some examples of real-world applications, refer to my research group's webpage[3].<br>In particular, your description should address the following concerns, among others:<br>• How does your proposed application exploit the desirable properties (e.g., predictive mean, predictive uncertainty, Bayesian, non-parametric) of GP models?<br>• In the context of your proposed application, what qualitative advantages (and/or limitations) do GP models provide, as compared to other regression models available in MSExcel or any other statistical tools/packages you're aware of?<br>• Does your proposed application have any important requirements (e.g., real time, data streaming in over time, big data) that can be satisfied by GP models or its more scalable variants (e.g., parallel/distributed/decentralized, online, or anytime sparse GP models)?<br>• How can the human users/experts in your proposed application interpret the outputs of the GP model and interact with it (if necessary)? More importantly, how do the outputs of the GP model help these human users/experts to plan and make decisions in your proposed application?[4] |
| 3% | **Technical approach**: Describe clearly and CONCISELY the technical details of the GP model or its more scalable variant that you are using. Doing so will allow you to demonstrate the theoretical rigor and the understanding of your research project work. Specifically, your description should address the following concerns, among others:<br>• How does the GP model you're using fit the requirements of your proposed application exactly?<br>• From your team discussions, what additional insights have you gathered about the GP model you're using, especially if its details are not covered during lecture?<br>• Can you propose some novel modifications/revisions (if any) of the GP model you're using to improve its performance (e.g., predictive accuracy, time efficiency) or to make it fit your proposed application better? |

[3] http://www.comp.nus.edu.sg/∼lowkh/research.html

[4] Note that in part 2 of this term project, you will be asked to automate some planning and decision-making pro-

| | |
|---|---|
| 3% | **Experimental evaluation**: Describe clearly and CONCISELY how you plan to empirically demonstrate the usefulness of the GP model in your proposed real-world application and empirically evaluate its performance (e.g., trade-off between predictive accuracy and time efficiency). In particular, you should consider the following guidelines:<br>• Describe your experimental setup clearly and carefully.<br>• What real-world dataset(s)[5] will you be using for training and testing?<br>• Are there more than one type of GP model/variant that can be used in your proposed application? Can other types of linear or non-linear regression models be used? If so, how do you plan to empirically compare their performance? What quantitative/empirical advantages (and/or limitations) do the GP model or its variant offer over the other evaluated models?<br>• You can consider other source codes of parallel/distributed and anytime GP models found online[6]. |
| 3% | **Wow! factor**: Basically, when I read your report, I should go like "Wow! Your research work is NOVEL, INTERESTING, and really cool!" In a technically sound and pleasantly surprising way, of course. Then, you'll get the marks in this category. |
| 1% | **Following instructions**: The instructions are provided on page 1. |

cesses based on the outputs of a GP model. For example, how can the GP model be used to inform the automated planning and decision-making strategies/algorithms on where to select and gather the most informative data/observations?

[5]Here are some links to real-world datasets:

- http://www.metoffice.gov.uk/hadobs/emslp/
- http://www.metoffice.gov.uk/hadobs/hadsst2/
- http://www.metoffice.gov.uk/hadobs/mohsst/
- http://www.epa.gov/airdata/
- http://www.esrl.noaa.gov/psd/data/gridded/
- http://select.cs.cmu.edu/data/index.html
- https://archive.ics.uci.edu/ml/datasets.html
- http://coastwatch.noaa.gov
- https://sites.google.com/site/goovaertspierre/pierregoovaertswebsite/download
- http://www2.hawaii.edu/∼matt/680/brooms_barn.txt

[6]Here are some links to the source codes:

- https://github.com/arikcj/pgpr
- https://github.com/qminh93/RVGP
- https://github.com/markvdw/GParML
- http://nbviewer.ipython.org/github/SheffieldML/notebook/blob/master/GPy/SVI.ipynb