Math 4432: Statistical Machine Learning	March 29, 2018

Project 2: Midterm

Instructor: Yuan Yao   *Due: April 12 11:59pm, 2018*

Requirement

1. Pick up ONE (or more if you like) favorite challenges *below*. If you would like to work on a different problem outside the candidates we proposed, please email course instructors about your proposal. Brave hearts for explorations will be encouraged!

2. Team work: we encourage you to form small team, up to THREE persons per group, to work on the same problem. Each team just submit ONE report, *with a clear remark on each person's contribution*. The report can be in the format of either Python (Jupyter) Notebooks with a detailed documentation, a *poster* such as

   https://github.com/yuany-pku/2017_math6380/blob/master/project1/DongLoXia_
   poster.pptx

   or a *technical report within 8 pages*, e.g. NIPS conference style

   https://nips.cc/Conferences/2016/PaperInformation/StyleFiles

3. In the report, show your proposed scientific questions to explore and main results with a careful analysis supporting the results toward answering your problems. Remember: scientific analysis and reasoning are more important than merely the performance tables. Separate source codes may be submitted through email as a .zip file, GitHub link, or as an appendix if it is not large. There is no restriction on the programming languages to use, but R or Python are recommended.

4. Submit your report by email or paper version no later than the deadline, to the following address (statml.hw@gmail.com) with Title: <u>Math4432: Project 2</u>. Late submissions may consume grades.

1   Kaggle contest classification: Predict survival on the Titanic

The following website contains the Kaggle contest on predicting survival (binary classification) on the Titanic:

  https://www.kaggle.com/c/titanic/

Register the Kaggle and join the contest by submitting your predictions. Report your methods and the corresponding scores (accuracy) on the leaderboard.

2   Kaggle contest regression: Predict house sales prices

The following website contains a Kaggle contest on predicting house sales prices (regression) using the Ames Housing dataset:

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/

It is aimed for practicing feature engineering, RFs, and gradient boosting etc. Register the Kaggle and join the contest by submitting your predictions. Report your methods and the corresponding scores (RMSE) on the leaderboard.