

Homework 3

Classification, regression, and other prediction model

Due Date: 23:59, December 22, Friday, 2017

TA: 蕭立品 vul1688@gmail.com

In this homework, the major task is prediction. There are a lot of prediction models, classification model predict labels, regression model predict real values, and some prediction model used in different types of data.

Dataset & Schema:

The same as the previous HW.

Tasks: What you need to do is using historical(96 hours) data to predict future(6 hour) data.

- How to produce dataset is shown as the picture below, this example consider Temperature as historical feature and NorthUsage as future target. You can change Temperature and NorthUsage to any data you like. For example, you can use historical NorthUsage to predict future NorthUsage, I'm just giving you an idea. You can also determine how far(1 hour, 2 hour, ...) the window will be slided each time.



2. **Randomly** split the dataset into training and testing part (70%, 30%).

dataset							
		
		
		
		
		
		
		
		
		
		
		

70% Training

30% Testing

3. Now, you have training dataset and testing dataset, use training dataset to train the following classification and regression model, and use testing dataset to evaluate them. Remember to do labeling before doing classification because the output of classification models is categories (labels).

- Classification:
 - K-Nearest-Neighbor
 - Naive Bayes
 - Random Forest
 - Support vector machine(SVC)
 - Others (bonus)
- Regression:
 - Bayesian regression
 - Decision tree regression
 - Support vector machine(SVR)
 - Others (bonus)

You can use default parameters in this task.

4. Tune the parameters of some models in Task.3. (bonus)
5. Try to define a classification (or regression) problem and solve it. (bonus)
I think this is a huge problem, don't worry about the score if you don't have time to do this task. =)

Report:

1. For task.1, briefly show the dataset you produce.

2. For task.2, briefly show me your training and testing dataset. I would like to see that you really split the data **randomly**.
3. For task.3, the performance(accuracy and computation time) of each model.
4. For task.4, the performance(accuracy and computation time) of models after tuned.
5. For task.5, clarify the problem you define, what method did you use, evaluation and comparison.