BT5152 Tutorial 4

AY 2018/19, Semester 1, Week 6

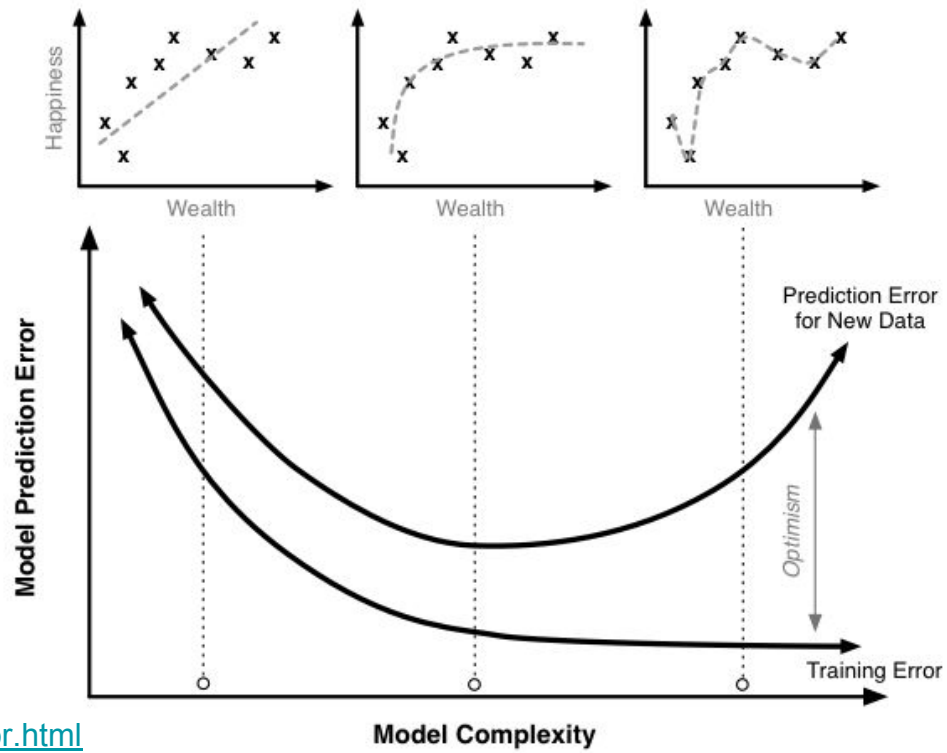
Lu Wei

Key Concepts Revision

Bias Variance Trade-off

- Overfitting: high bias or high variance?
- Underfitting: high bias or high variance?
- Can a model has both high bias and high variance?

No



Graph from: http://scott.fortmann-roe.com/docs/MeasuringError.html

Bagging

sample with replacement

generate multiple samples with the same data

train data is limited

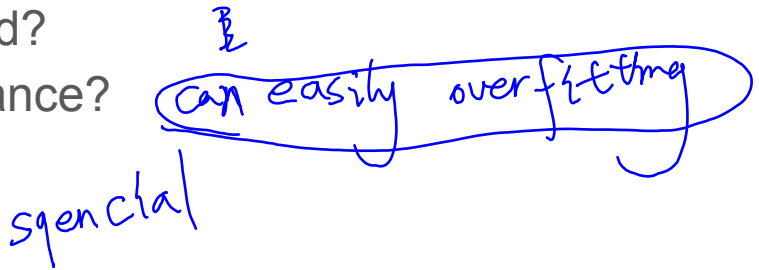
- Bootstrapping + averaging (for regression) / voting (for classification)
- Helps reducing bias or variance?
- Bagging vs. cross validation
- Random Forest vs. Tree Bagging

bagging used in decision tree

choose rows & columns

Boosting

- Weighted bootstrapping + averaging (for regression) / voting (for classification)
- How are weights determined?
- Helps reducing bias or variance?
- Boosting vs. Bagging



Stacking

- Multiple primary models + 1 meta model
- Helps reducing bias or variance?
- How to pick which primary models to include?

try all the algorithms and aggregating function

hard to tune

Ensemble Learning: Pros & Cons

- Well-performing, robust models
- Poor interpretability
- Takes time to build & tune

Assignment 2

- Due 2018-10-02 5:59pm
- Name your files using your student number: e.g. A0123456X.R, A0123456X.html, A0123456X.csv
- Do not zip
- Q2, Q3: set.seed(42) at the start of your solution code
- Q3: top 10% submissions will receive bonus +1 mark, bottom 10% submissions will receive penalty -1 mark

Tutorial Exercises:

RStudio > Console:

```
# install.packages("swirl")
library(swirl)
# delete_progress('your name')
install_course_github('weilu', 'BT5152', multi=TRUE)
swirl()
```

1: Bagging
2: Boosting
3: Stacking