# Study on tree-based methods. MATH 6380 project 2

Chenyang, DONG Tsz Cheung, LO Jiacheng, XIA

April 22, 2017

## Outline

- Introduction
- 2 American Crime Dataset
- Staggle 1: ComboDrug
- Maggle 2: Binray Drug
- 6 Analysis and Conclusion

- Introduction
- 2 American Crime Dataset
- Kaggle 1: ComboDrug
- 4 Kaggle 2: Binray Drug
- 5 Analysis and Conclusion





Why did we choose tree-based methods?

• We went through several methods.



- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.

- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.
- There are yet many improvement methods.

- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.
- There are yet many improvement methods.
- Studied the method on 3 datasets.



- Introduction
- 2 American Crime Dataset
- 3 Kaggle 1: ComboDrug
- 4 Kaggle 2: Binray Drug
- Analysis and Conclusion

#### The dataset

This dataset and the preprocessing are the same as project 1.

#### The dataset

This dataset and the preprocessing are the same as project 1.

#### Goal for this dataset

Do straightforward analysis and compare with Lasso (PJ 1).

#### What we found

Simple regression tree out-perform Lasso; bagging, random forest and boosting even better.

- Introduction
- 2 American Crime Dataset
- Kaggle 1: ComboDrug
- 4 Kaggle 2: Binray Drug
- Analysis and Conclusion

- Introduction
- 2 American Crime Dataset
- Kaggle 1: ComboDrug
- Maggle 2: Binray Drug
- 5 Analysis and Conclusion

# Models and Kaggle Results

Dummy1



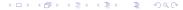
# Models and Kaggle Results

Dummy1 Dummy2



# Variable Importance

Dummy



# Compared with p-value Selection

Dummy



## Some Conclusions

We came up with some conclusions (inferences):

## Some Conclusions

We came up with some conclusions (inferences):

dummy1



## Some Conclusions

We came up with some conclusions (inferences):

- dummy1
- dummy2

- Introduction
- 2 American Crime Dataset
- 3 Kaggle 1: ComboDrug
- 4 Kaggle 2: Binray Drug
- 6 Analysis and Conclusion