CyberSciTech 2021

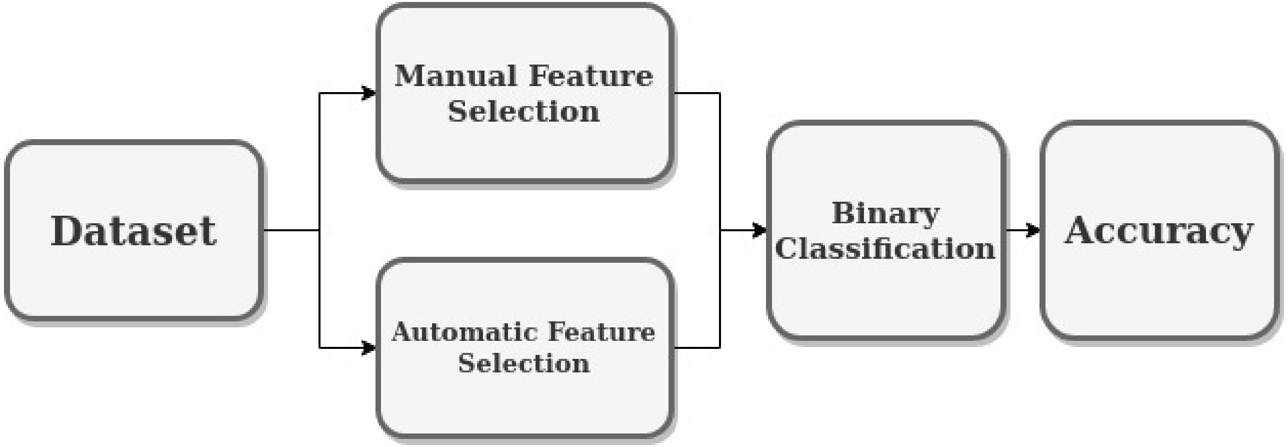
Ransomware Detection Using Binary Classification

Kazi Samiul Kader, Md. Tareque Hasan, Md. Shohrab Hossain, Husnu S. Narman

Motivation

- Ransomware attacks are on the rise
- It is computationally infeasible to reverse such attacks
- Signature based detection is not enough as ransomwares evolve
- Dynamic detection is more effective
- Machine learning can be used in dynamic detection of ransomware
- Our *objective* is to detect ransomware accurately using binary classification algorithms

Overall View


Dataset

Manual Feature Selection

Automatic Feature Selection

Binary Classification

Accuracy

Our Dataset

- 1524 rows
  - 582 ransomwares
  - 942 good applications
- 30970 columns (features)
- Features are different operations performed at installation by an application or ransomware

Feature Selection

- Too many columns (features), too little rows
- 2 type of feature selection
    - Manual (category wise)
    - Automatic (chi-square test)

Manual Feature Selection Categories

- API invocation
- Extension of dropped files
- Registry key operations
- File operations
- Extension of the files involved in file operations
- File directory operations
- Embedded strings

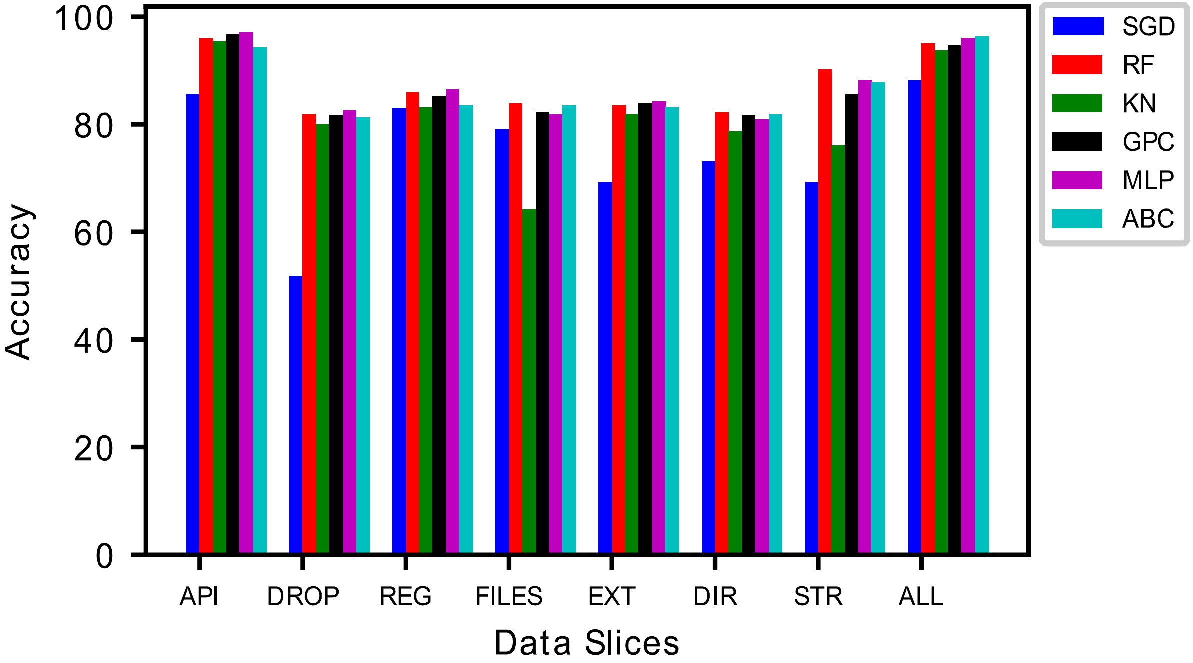
Automatic Feature Selection

- K Best features using the chi-square test
- K = 100 – 300

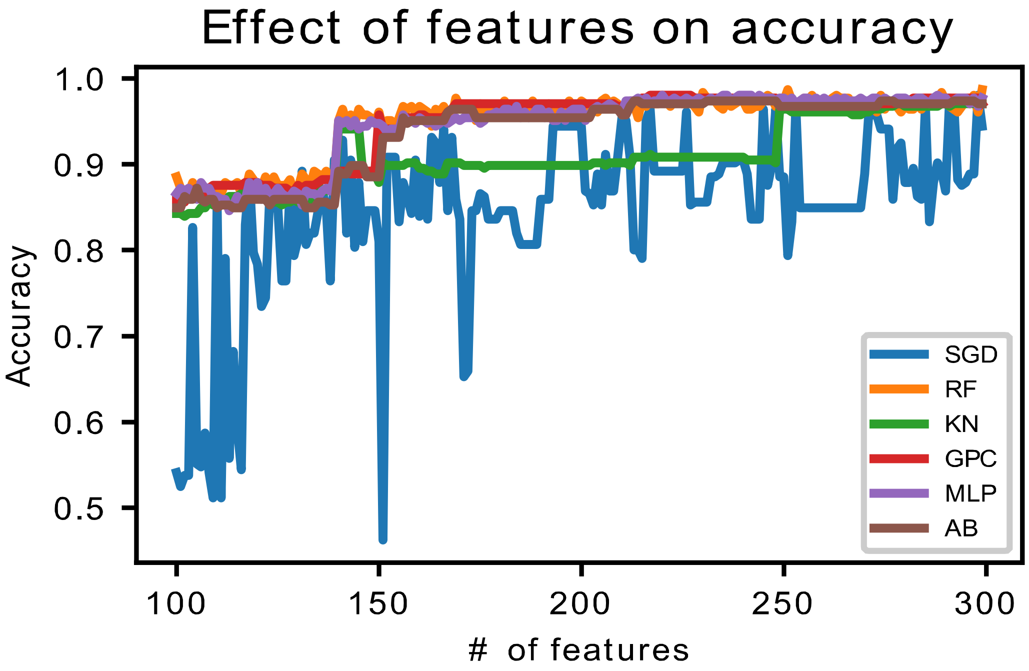
Binary Classifiers

- Stochastic Gradient Descent Classifier
- Random Forest Classifier
- K Neighbors Classifier
- Gaussian Process Classifier
- Multi-Layer Perceptron Classifier
- Ada Boost Classifier

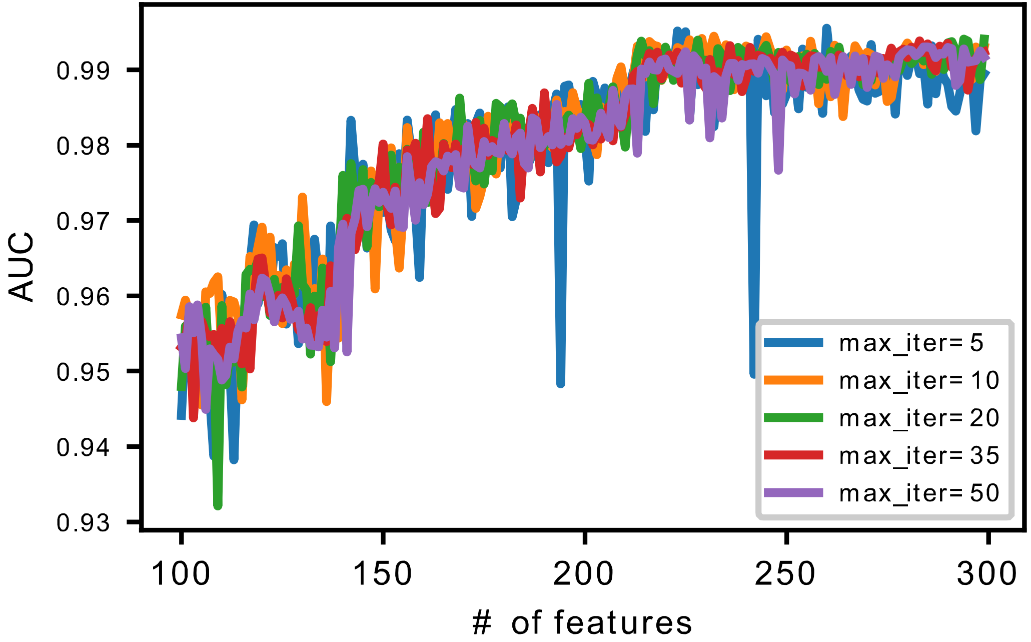
Manual Selection Results



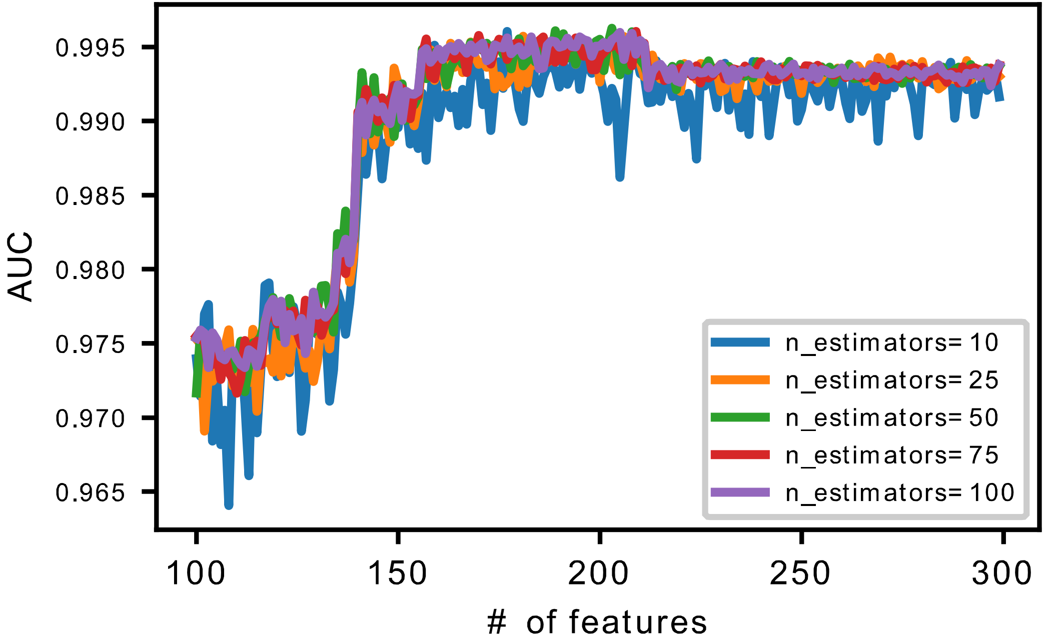
Automatic Selection Results



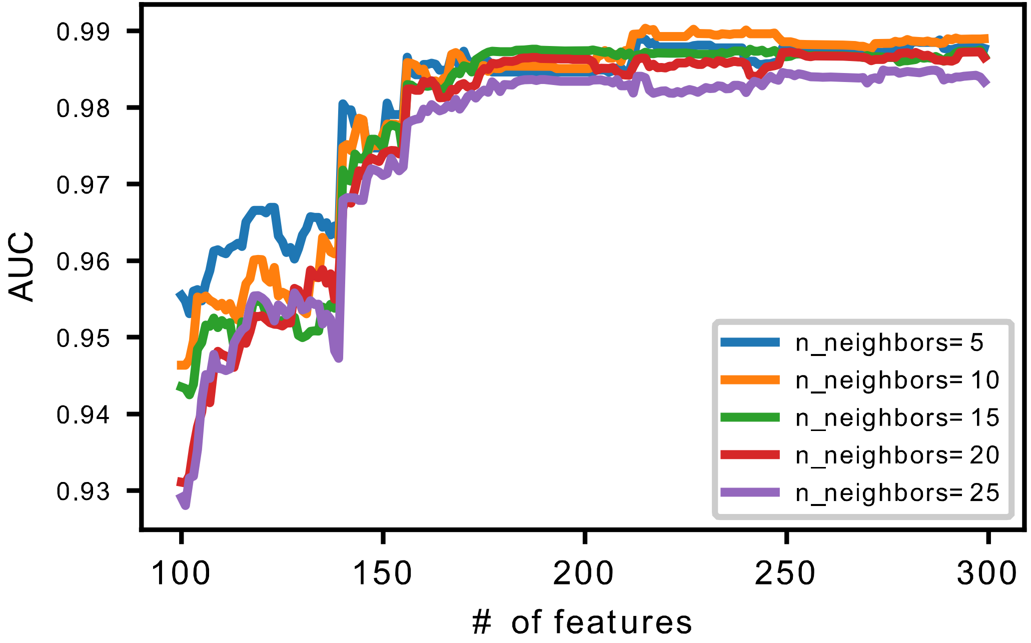
SGD Parameter Sensitivity



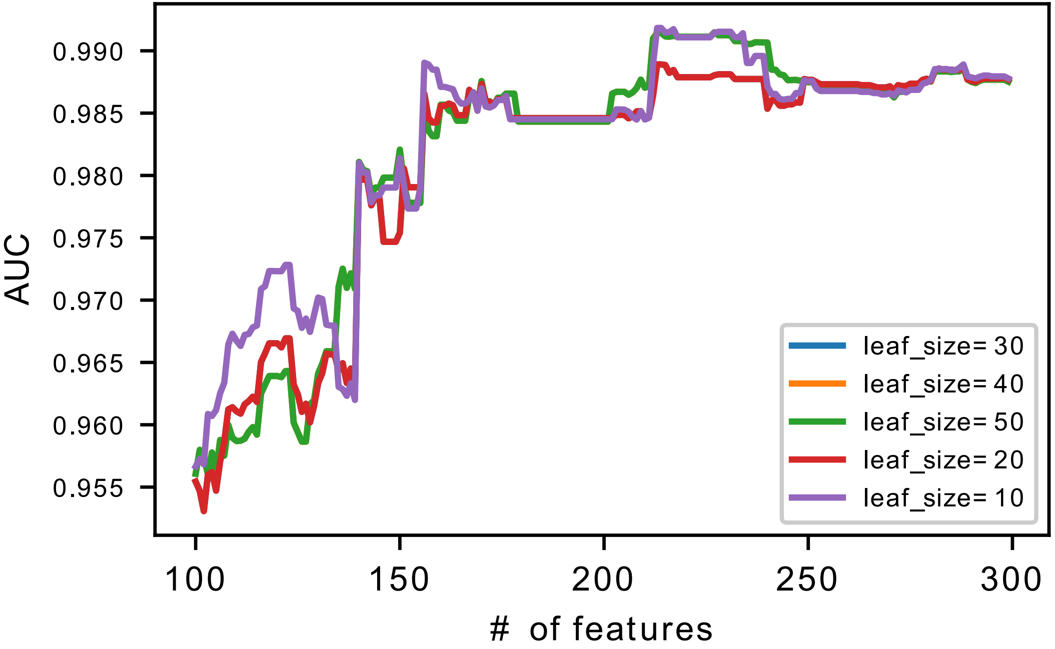
RF Parameter Sensitivity



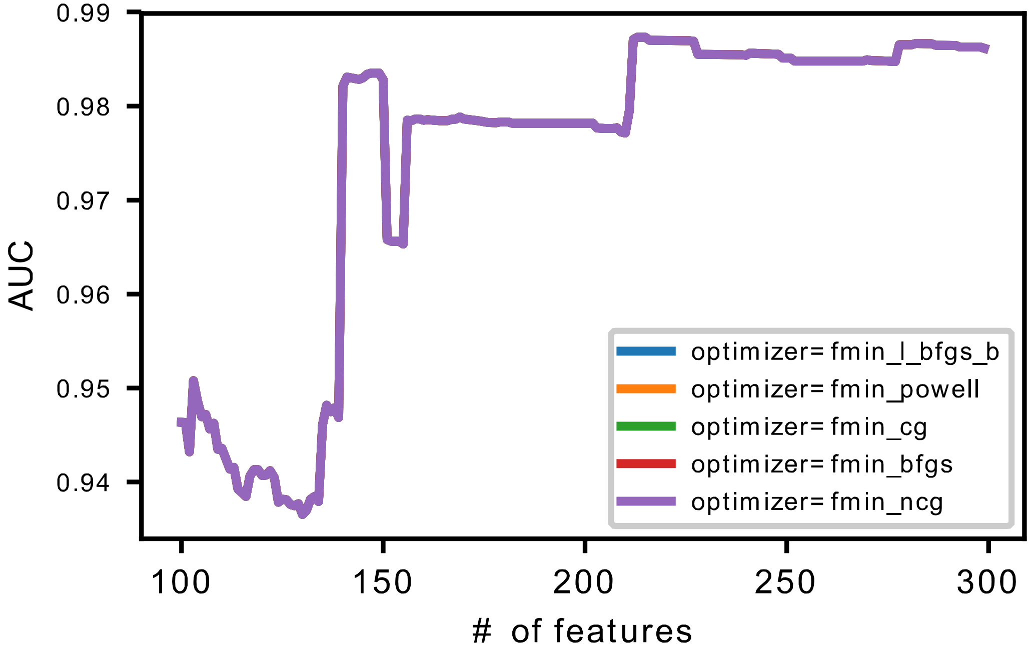
KN Parameter Sensitivity



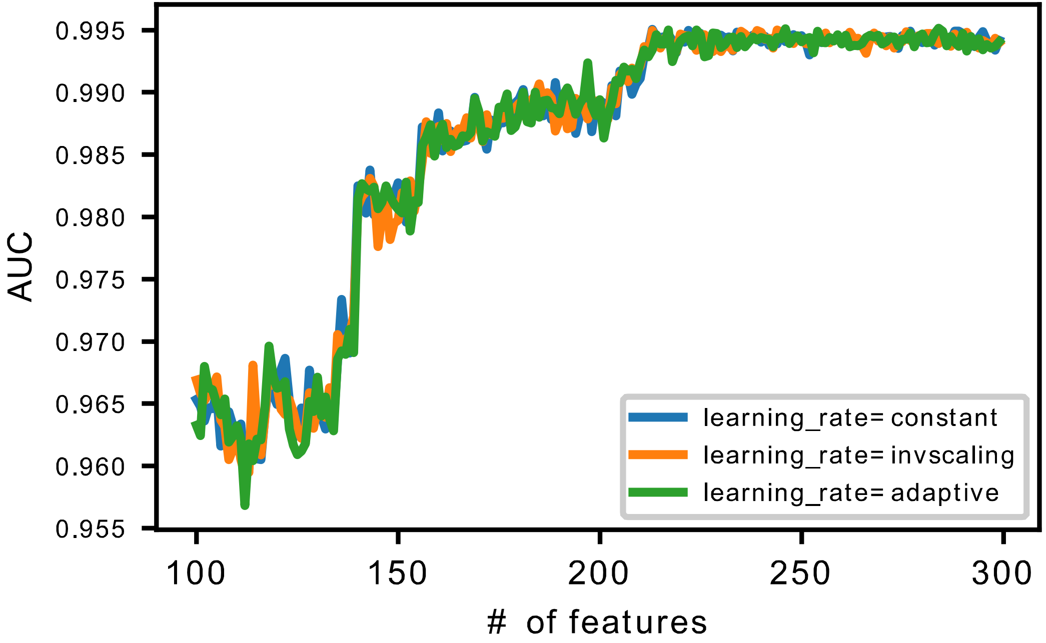
KN Parameter Sensitivity



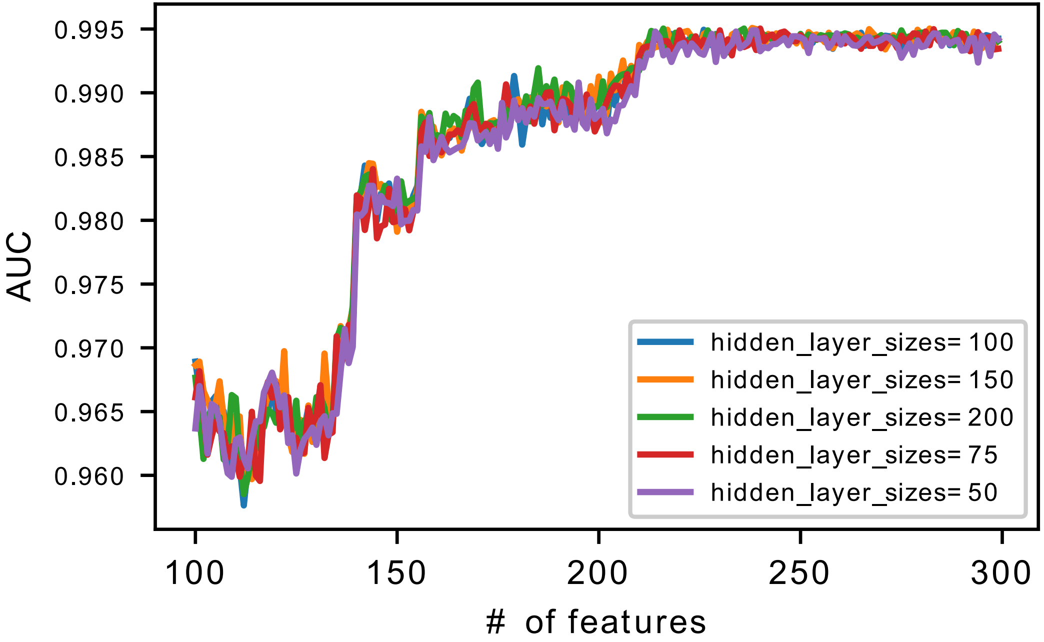
GPC Parameter Sensitivity



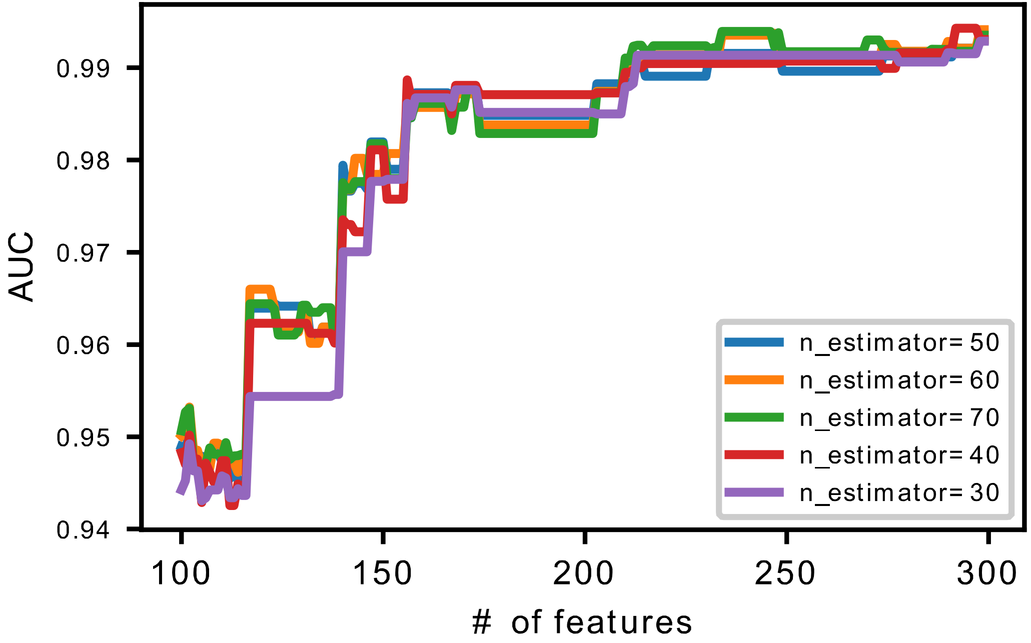
MLP Parameter Sensitivity



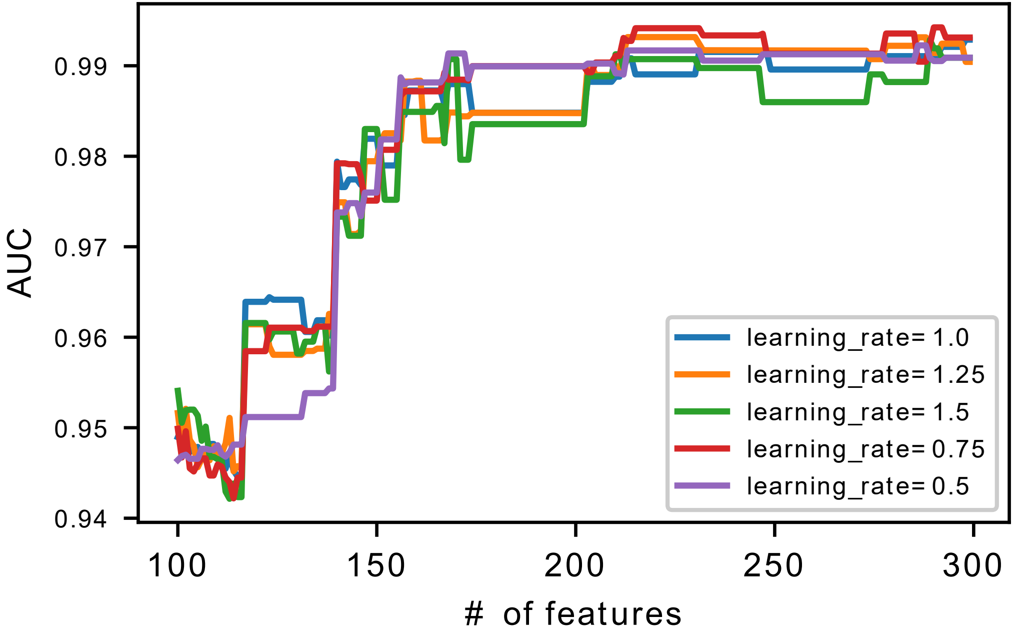
MLP Parameter Sensitivity



Ada Boost Parameter Sensitivity



Ada Boost Parameter Sensitivity



Conclusion

- Effectiveness of simple machine learning algorithms
- Application of deep learning algorithms in future

Thank You