

# Statistics 315B Homework 2

Spring 2017

May 17, 2017

## 1

This answer is referred to in [1]’s *random subsampling method* and [2].

### 1.a Advantages

- Free from curse of dimensionality
- The algorithm is scalable.(A parallel algorithm)
- It does not get hit by bootstrapping or bagging in generalization.
- Effectively eliminating the feature redundancy inherent in the data.
- Effectively reducing the correlation between estimators.

### 1.b Disadvantages

- The original random subspace method did not emphasize the importance of selecting a base learner that should generalize well.

### 1.c

Bagging or bootstrapping is possible. However, some papers and work[3][4] have also introduced parametric or nonparametric heuristic procedures or manual interventions.

## 2

### References

- [1] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.

- [2] Xiaoye Li and Hongyu Zhao. Weighted random subspace method for high dimensional data classification. *Statistics and its Interface*, 2(2):153, 2009.
- [3] Seymour Shlien. Multiple binary decision tree classifiers. *Pattern Recognition*, 23(7):757–763, 1990.
- [4] Seymour Shlien. Nonparametric classification using matched binary decision trees. *Pattern Recognition Letters*, 13(2):83–87, 1992.