

# Statistical Machine Learning II Project Proposal

Ze Yang, Zhengyang Qi, Yundong Liu, Yuze Liu

April 9, 2018

Our final project will be based on the paper “Understanding Black-Box Predictions via Influence Functions” by Pang Wei Koh and Percy Liang[1], which discusses how influence function can trace the models prediction through the learning algorithm and the training data. For now, we decide to divide the projects into three parts. First of all, we will figure out the mathematical formula behind influence function. Secondly, we will study and implement the two techniques in efficiently calculating influence function discussed in the paper. Last but not least, we will apply the influence functions to other algorithms that we have learned in class and that are not discussed in the paper (e.g. Ridge Regression and Trees). Through this project, we intend to have a relatively thorough understanding of how influence functions can give us insights on training data.

## References

- [1] Pang Wei Koh, Percy Liang *Understanding Black-box Predictions via Influence Functions*. International Conference on Machine Learning, 2017.