

CSC 240/440 - 2017 Spring: Course Project

1 Description

The course project of this project is the competition “House Prices: Advanced Regression Techniques” on kaggle <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>. See the link for an overview description.

This competition is about predicting the the final price of each home, with 79 features describing (almost) every aspect of residential homes. A very good detailed analysis visualization of some features can be found at <https://www.kaggle.com/pmarcelino/house-prices-advanced-regression-techniques/comprehensive-data-exploration-with-python/run/879421>. (Here you may also learn how to read in the data using pandas).

In the kaggle website, you can use the “kernels” or “discussion” board to learn from others, or make discussion with professors, TAs, or other students. However, you are **not allowed** to directly use the solutions/examples provided online (or from other group). If your final submission (code, document) directly copy (or make simple modification) from there, we see it as cheating (see Academic Honesty Policy in the syllabus).

2 Requirements

This is a group project, each group can have up to 3 members. We encourage you to form team with diversity. One bonus point will be given to the team which has members from different majors or contains both undergraduate students and graduate students.

You need sign up an account in kaggle, and make a team name (so your submitted result will be shown with this name in the leaderboard)¹.

In the end, you need submit your code and a report. The report should include the details of your implementation, the methodology of your algorithm. We encourage you to organize the format as sections, i.e. Introduction (the reason/motivation why you choose such algorithms), Method (How you do to get the result), Experiment Result (What result you get e.g. rank in the leaderboard, and also some analyses).

You also need to make a presentation in class, which takes 20% of the project points.

We encourage you to start early, firstly try to used some simple features (e.g. numeric features) and models (e.g. linear model), then try to improve your performance iteration by iteration.

¹in terms of group

3 Grading Scheme

Your grade will significantly depends the final performance of your algorithm, while the presentation will occupy 20% points. Each team is required to submit a final project. The team achieving top 10 in the leader board will receive 5 bound points. The team achieving top 1 in our class will receive 3 bound points.

Since March 23, in every Thursday's class (before the end of this semester), we will select top **3** teams to receive 1 bonus point. But the team in top 3 cannot earn this bound point if its performance is not strictly better than the one it got in last week. You also need upload your team name to blackboard so we can know your performance/rank on the leaderboard.