MoonBoard Grade Classification - May 2024
Group Members: Prabhjot Singh, Larsen Linov, Gautam Prakriya, and Adrian Batista Planas

Project Overview

The MoonBoard is a standardized climbing wall with a fixed set of climbing holds that are set at a specific position and orientation on the board. The standardization allows for users to climb the same wall as friends and other MoonBoard users. There are a number of versions of the MoonBoard; we focus on the original (2016) version.

The Moonboard climbing wall also comes with an app that curates user generated climbing routes/problems. Each problem is assigned a subjective grade that represents its difficulty. Since the moonboard is standardized and multiple users attempt the same problems, there is often a consensus to the grades assigned to moonboard problems making them somewhat objective. There are a number of grading systems for climbing routes. The labels in our data use the french grading system. The routes are assigned grades from the set '6B', '6B+', '6C', '6C+', ….., '8B', '8B+'; '6B' being the easiest and '8B+' being the hardest.

The aim of this project will be to develop a model(s) to classify/predict the grade of a given MoonBoard problem.

Stakeholders involved in this project include:

1) Moon Climbing - the company selling the MoonBoard, and
2) Climbers and MoonBoard users.

Key Performance Indicators (KPIs) that we used to evaluate our models included:

- Exact accuracy - Fraction of predictions that are exactly correct
- One-off accuracy - Fraction of predictions that are at most one grade off the true grade.

Data Cleaning and Preprocessing

For the sake of simplicity and since the 2016 MoonBoard model is the oldest and has the most repeated problems, we decided to restrict our modeling to only the problems possible with the 2016 MoonBoard. To clean the data, problems were removed if they had fewer than 2 repeats, a user rating of 3 or below, and/or there was an inaccurate hold (e.g., A6 is not a hold in the 2016 model, so problems with A6 were removed). Preprocessing involved converting grades (e.g., 6C+, 7A, 8B) to integers (i.e., 1-13).

Modeling Approaches

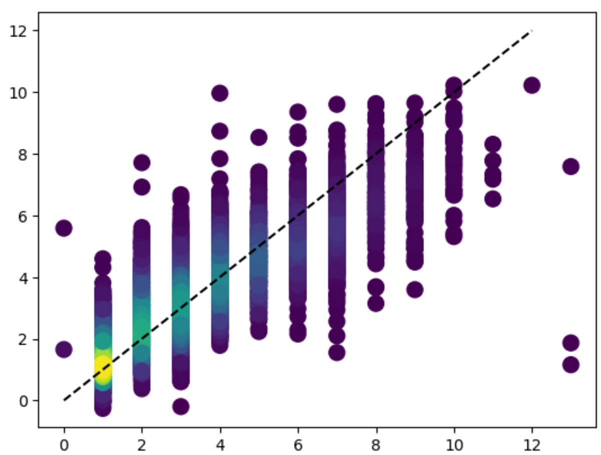
First, as a baseline model, we applied linear and logistic/softmax regression using the presence of each hold (encoded as 0 or 1) as our feature set.

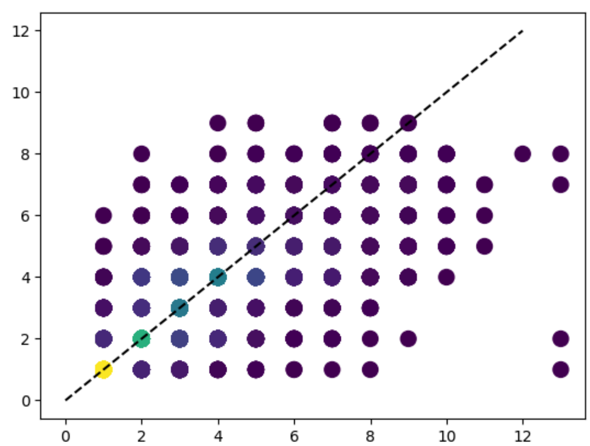
Our predictions come from adding synthetic features and using the XGBoost regressor and neural networks (via PyTorch) with the enlarged feature set. Our synthetic features are designed to take into account the geometric layout of the holds on the board. We included standard summary statistics for a collection of numerical values as well as, for example, the size of the maximal vertical gap between holds for each problem. In addition, we created an algorithmic estimate of a likely order for each of the holds to be used and thereby obtained a large set of bigram variables, which were run through UMAP to reduce them to 15 dimensions. However, we didn't include those 15 features in our final model because we found that they did not improve accuracy.

In our modeling, the neural network treated the problem as a multiclass classification problem with one class for each grade, ignoring the natural order of the grades. The neural network was set to optimize the cross-entropy of the classification. The XGBoost regressor, like our baseline models, used the naive numerical system of assigning integral values for each grade, with adjacent grades spaced 1 apart from each other. It then gave continuous estimates for each problem, which can be rounded to the nearest integer to obtain predicted classes. The XGBoost regressor was set to optimize the mean squared error of the numerical predictions (before rounding).

Results

After cross-validation, the hyperparameters for the XGBoost regressor were set to 300 estimators and a maximum tree depth of 3. The final mean squared error for the test set was 1.37, with an accuracy of 40% for exact grade matches and 83% for grade matches within one step of the "true" value. The neural network achieved a cross-entropy of 1.56, an exact match rate of 51%, and near-match rate of 77%. Confusion graphs for each model are shown below: On the left is the XGBoost regressor, and on the right is the neural network. In each case, the horizontal axis represents the numerical index for the "true" grade (0 = 6B, 1 = 6B+, …) and the vertical axis is the predicted grade.





For comparison, our baseline models obtained accuracies of 30% and 40%, respectively, for exact grade matches and 75% each for near-matches. Human climbers are said to have an accuracy of 45% for exact matches and 85% for near-matches, although those numbers might have been obtained with different datasets.

Limitation and Future Work

Two prevalent limitations to our project include conflicting classification of problems and the limited range of grades we included in our model. Since various users provide subjective grades for each problem and any MoonBoard user can create a problem and assign a grade, there may be inconsistencies in the grades for the same problem used to train the model.

However, since the grade itself is a subjective factor to begin with, and we are asking the model to classify into subjective, man-made categories, it is natural to see disagreement and contradictions when classifying the problems. Another limitation is that the MoonBoard only features problems between grades ranging from 6B - 8B+.

Future analysis could involve expanding the database of problems included. This would include expanding the grade range of problems beyond 6B-8B+, including problems from the 2017, 2019, and 2024 MoonBoard models, and adjusting the weight of grades based on whether the board is setup at 25° or 40°.