MARCH MACHINE LEARNNG MANIA 2021-NCAAW SPREAD

In stage one of this two-stage competition, participants will build and test their models against previous tournaments. In the second stage, participants will predict the point spreads of the 2021 tournament. We don’t need to participate in the first stage to enter the second. The first stage exists to incentivize model building and provide a means to score predictions. The real competition is forecasting the 2021 results.

**To Predict**

**Stage 1** - You should submit predicted point spread for every possible matchup in the past 5 NCAAW® tournaments (seasons 2015-2019).

**Stage 2** - You should submit predicted point spread for every possible matchup before the 2021 tournament begins.

**Dataset**

**WNCAATourneyDetailedResults.csv**

This file provides team-level box scores for many NCAA® tournaments, starting with the 2010 season. All games listed in the WNCAATourneyCompactResults file since the 2010 season should exactly be present in the WNCAATourneyDetailedResults file.

submit\_stage1\_elo\_tuned.csv

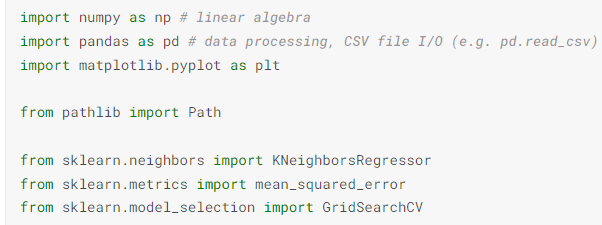
elo\_2021-04-05.csv

elo\_calibrated\_margin\_submitted.csv

2021NCAAWTourneyMarginResults.csv

elo\_calibrated\_margin\_fix\_season\_merge.csv

Step 1: Importing the libraries



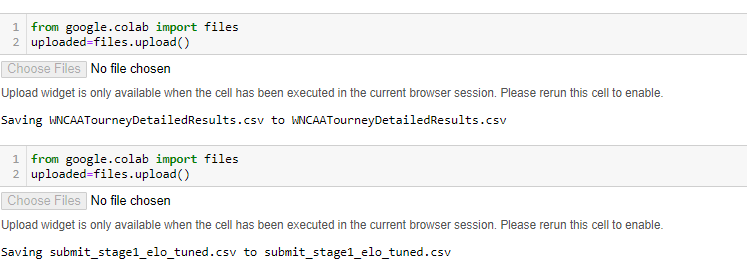
Regression based on k-nearest neighbors.The target is predicted by local interpolation of the targets associated of the nearest neighbors in the training set.

Mean squared error regression loss.

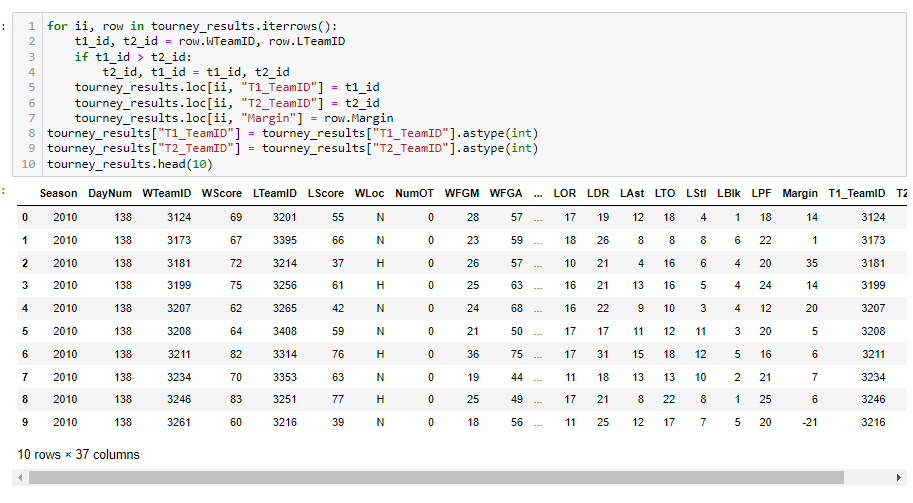
GridSearchCV implements a “fit” and a “score” method. It also implements “predict”, “predict\_proba”, “decision\_function”, “transform” and “inverse\_transform” if they are implemented in the estimator used.

The parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid.

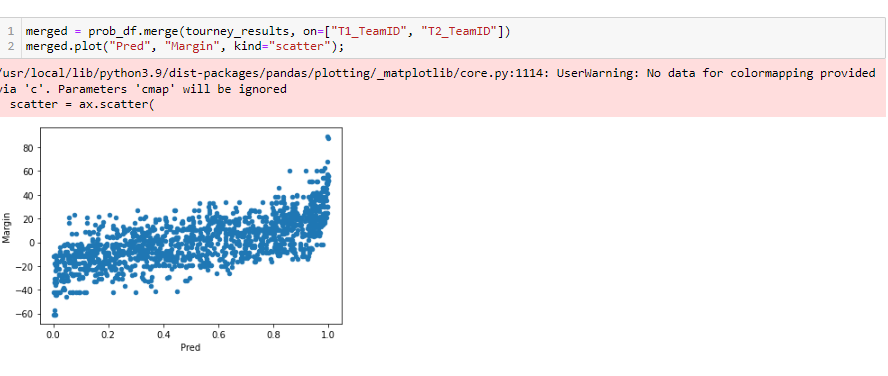
**Step 2: Loading the data**



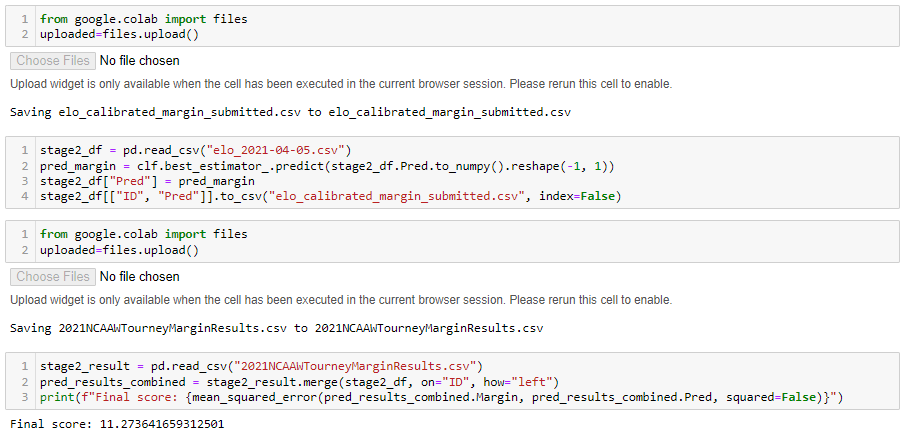
**Step 3: add T1/2\_TeamID to the tournament results.**



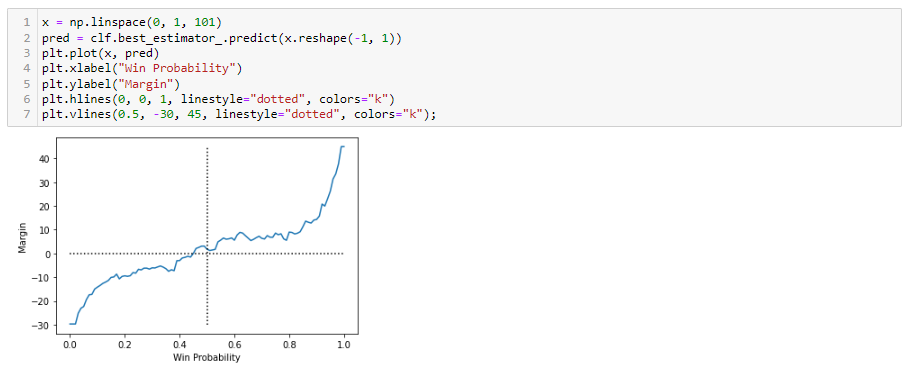
**Step 4: stage1 submission with the tournament results.**



**Step 5: Checking the Final Score**



**Step 6: Plotting the Model**



There is a lot of noise in these competitions and to some degree I got lucky with the asymmetry. Although there is just a chance asymmetry in the data as well. Alphabetically lower teams in the NCAAW are just a little stronger than alphabetically higher teams.