

## Lab 2: Logistic regression using scikit-learn

The *nbaallelo\_log.csv* file contains data on 126314 NBA games from 1947 to 2015. The dataset includes the features *pts*, *elo\_i*, *win\_equiv*, and *game\_result*. Using the csv file *nbaallelo\_log.csv* and scikit-learn's *LogisticRegression()* function, construct a logistic regression model to classify whether a team will win or lose a game based on the team's *elo\_i* score.

Step 1: Create a binary feature win for *game\_result* with 0 for L and 1 for W

Step 2: Use the *LogisticRegression()* function to construct a logistic regression model with *win* as the target and *elo\_i* as the predictor

Step 3: Print the weights and intercept of the fitted model

Step 4: Find the proportion of instances correctly classified

Note: Use *ravel()* from numpy to flatten the second argument of *LogisticRegression.fit()* into a 1-D array.

Ex: If the program uses the file *nbaallelo\_small.csv*, which contains 100 instances, the output is:

```
w1: [[3.64194406e-06]]
w0: [-2.80257471e-09]
0.5
```

**Submission:** Submit the completed *lab2.py* file and *written.pdf* where you show your output.