Predicting the winning side of DotA2
Kuangyan Song, Tianyi Zhang, Chao Ma

INTRODUCTION
- A DotA 2 game has ten players with five on each side. Each player decide for himself which hero to use.
- Other than the experience level of the player. Lineups on either side has an influence on the result of the game.
- Certain combination of heroes are complimentary while some others are restrictive which might have an impact on the game.

METHODOLOGY
Baseline
- Without taking the experience level of the players into account, consider the lineups as the only feature. Using logistic regression to fit the data.

\[
\phi(X) = \begin{cases} 
X_i & \text{if player of radiant side plays hero } i \\
0 & \text{otherwise}
\end{cases}
\]

\[
Y_i = \begin{cases} 
1 & \text{if radiant side won} \\
0 & \text{if dire side won}
\end{cases}
\]

\[
\min_w \frac{1}{2} w^T w + C \sum_{i=1}^{m} \log(1 + e^{-y_i \phi(x_i) w + c}) + 1
\]

DATA COLLECTION
- Match and player data are retrieved via APIs provided by the game developer.
- We only consider those games modes that is “All Pick”, “Ranked All Pick” and “Random Draft”. Moreover, those games that has absent players or has zero kills throughout the game are considered to be invalid.

FEATURE SELECTION
- Without taking the experience level of the players into account, consider the lineups as the only feature. Using logistic regression to fit the data.
- Both the training error and test error are high. Most of the data points are very close to the decision boundary, we might have too many features. The assumption that all the players are equal is also questionable.

Adding features for hero combo
- Before selecting features, we took the interaction of heroes into account. To simplify the problem, we hand picked fifty known significant hero combination to observe as features, now we have 324 features.

FUTURE WORK
According to our work, we can see that predicting the winning side based on hero selection is not enough. Future work should take the player into account to make the prediction more accurate. Moreover, the model might need to be changed to get a better result.