

Exploration of Baseball Pitching Data

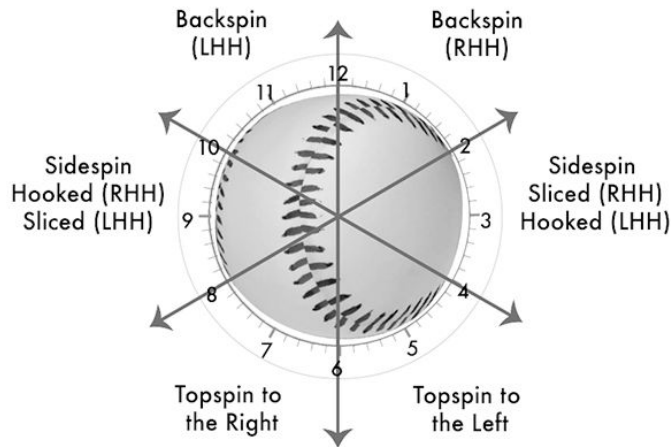
By: Alexander Donald and Gabriel Levy

Features and Outcome

- Outcome: pitch type (fastball, slider, changeup, curveball, cutter, splitter, sinker)
- Features
 - Spin rate: how many rpms the ball is spinning at
 - Spin direction: axis on which the ball is spinning
 - Spin efficiency/active spin: percentage of spin that works to move the ball in a certain direction
 - Movement ratio (maybe): ratio of induced vertical break and horizontal break

Data Preprocessing

- Data from baseballsavant.mlb.com
- Convert spin direction (given in hours and minutes deviated from 12 o'clock, positive and negative) to absolute value decimal number with hours to the left of decimal point and minutes/60 to right
- Normalize all features to have mean of 0 and standard deviation of 1



Model

- Only 3 features - KNN would be good fit
- Could also look into neural networks or support vector machines
 - More than just two outcomes, so would need to use multiclass versions of models

Ethical concerns

- No real moral concerns, since it's just baseball
- Some types of pitchers will be underrepresented in the data set and could vary significantly from the training data (sidearm and submarine pitchers)
 - These pitchers often have very different spin axes

