Modeling Pitch Tunneling in Baseball with Deep Learning

Jacob Lapp

jl372@rice.edu

Objective:

- Train a neural network that can:
  - Predict change in pitch values of consecutive pitches
  - Identify the most effective two-pitch sequences

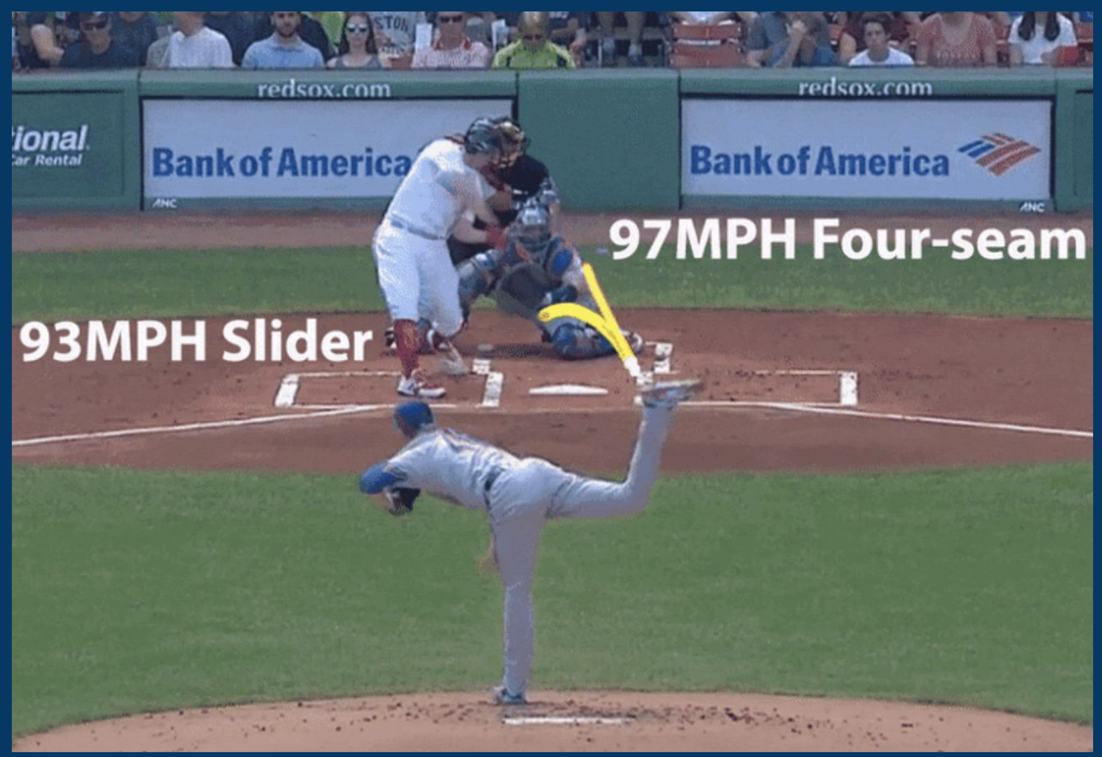


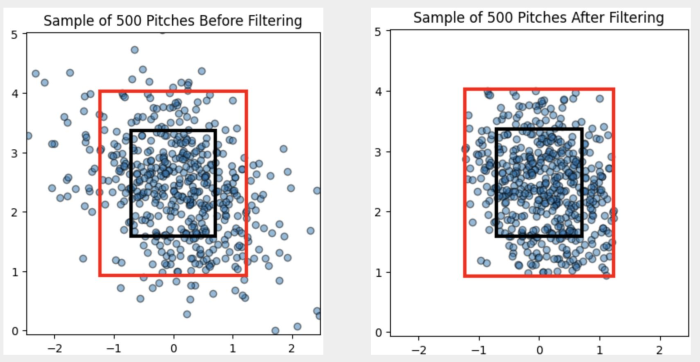
Image Courtesy of: https://alexgravellebusiness.medium.com/a-study-in-pitch-tunneling-bfb7b8363394

Introduction:

- Phenomenon in baseball known as **'pitch tunneling'**
  - Successful 'tunneling':
    - Pitch 1 and pitch 2 have identical initial trajectories
    - 1 of 2 pitches diverges (or not)
    - Confuses hitters, harder to identify which pitch is which
    - Decreases hitters' decision time

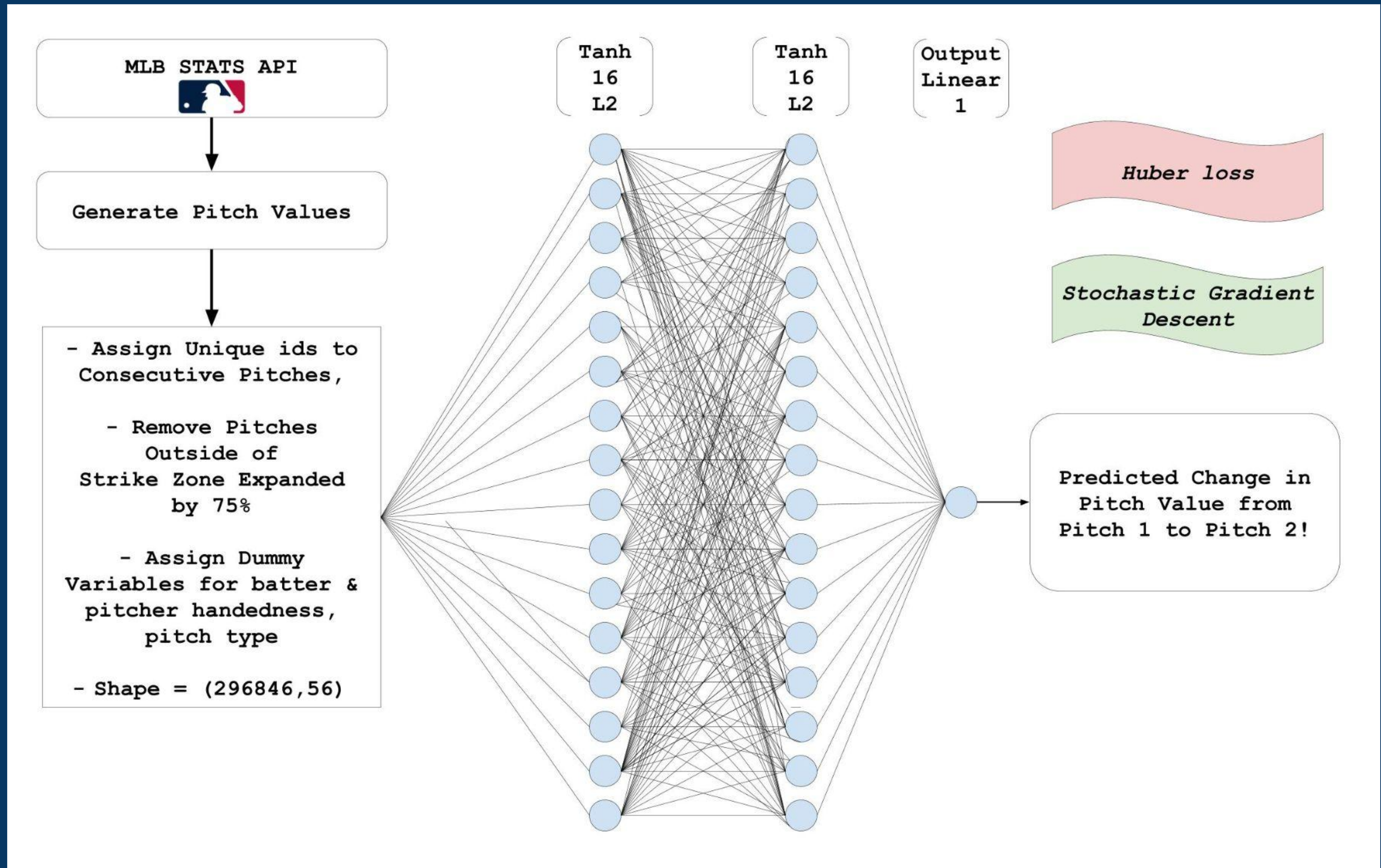
Preprocessing:

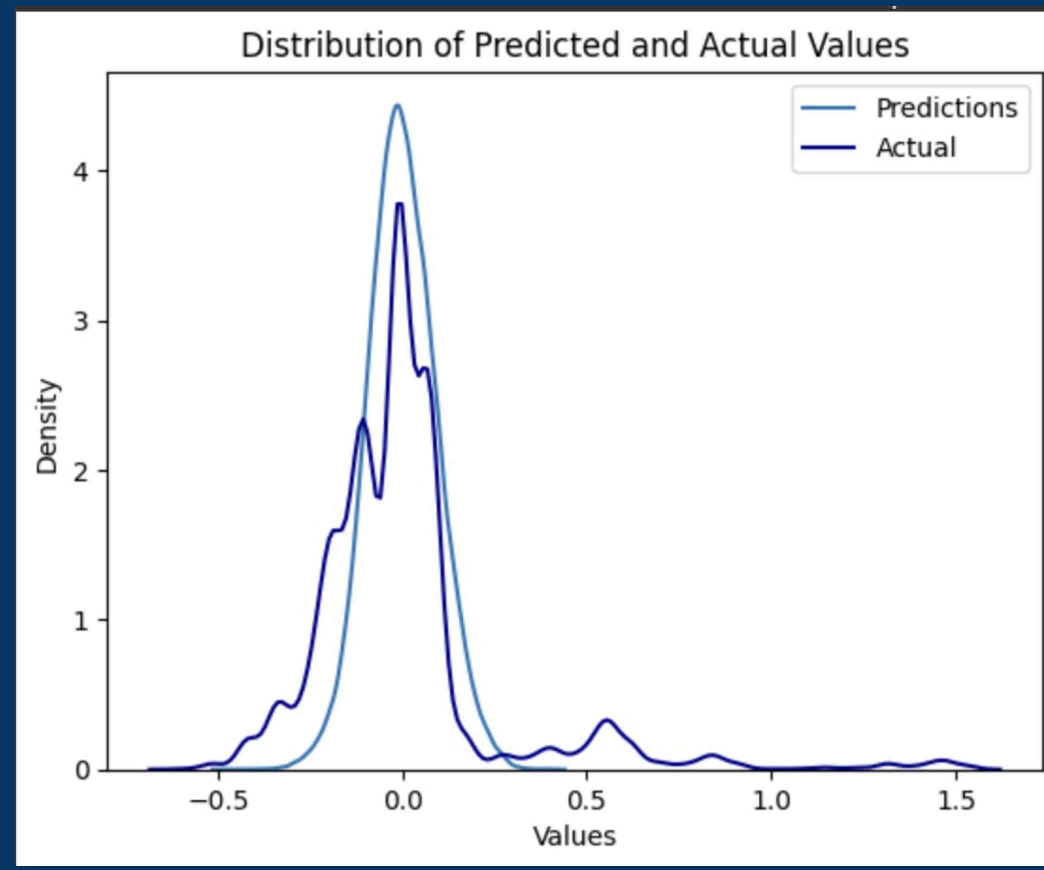
- Assign pitch value to every pitch via pitch outcome model
- Assign unique ID to each pitch appearing in a consecutive pair
- Filter out pitches that lie outside of strike zone enlarged by 75%
- Calculate differences across pairs, OHE categ.



Methods:

- Split into train and test sets
- Target variable:
  - Difference in pitch value across pair
- Train Neural Network
  - Loss: *Huber*
  - Stochastic Gradient Descent
  - Hidden Layers: 16 Neurons (x2)
  - Activation: *tanh, L2* Regularization
  - Final Activation: *Linear*
- Analyze predictions:
  - Permutation Feature Importance
  - Gradient Ascent on input features





Results:

- Test predictions follow roughly same distribution as actual test data
- Permutation Feature Importance:
  - First pitch *Fastball*
  - Second pitch *Cutter*
  - First pitch *Sweeper*
  - Second pitch *Sweeper*
- Gradient Ascent Input Features:
  - Diff. z-acceleration
  - Diff. release point z-coord
  - Second pitch *Fastball*
  - First pitch *Sinker*

Discussion:

- Can use model to project how pitchers' repertoires will perform
- *Fastballs* → *Sweeper* / *Cutters* (& vice-versa) seem to perform well
- Next version:
  - Differences across pairs conditioned on initial trajectory similarity