PREDICTING THE PERFORMANCE OF MLB PITCHERS WITH NEURAL NETWORKS & BAYESIAN INFERENCE

OMAR RAHEEM

Problem Statement

Goal One:

One expand on the scientific paper entitled: “Ball Speed and Release Consistency Predict Pitching Success in Major League Baseball” by David Whiteside , Douglas N Martini, Ronald F Zernicke, Grant C Goulet. By using a neural network and an array of features contrary to those used by the authors I aim to expand on their study

Goal Two:

To use Bayesian Inference to predict 2019 season ERA of 3 Hall of Fame pitchers.

Terms to Know

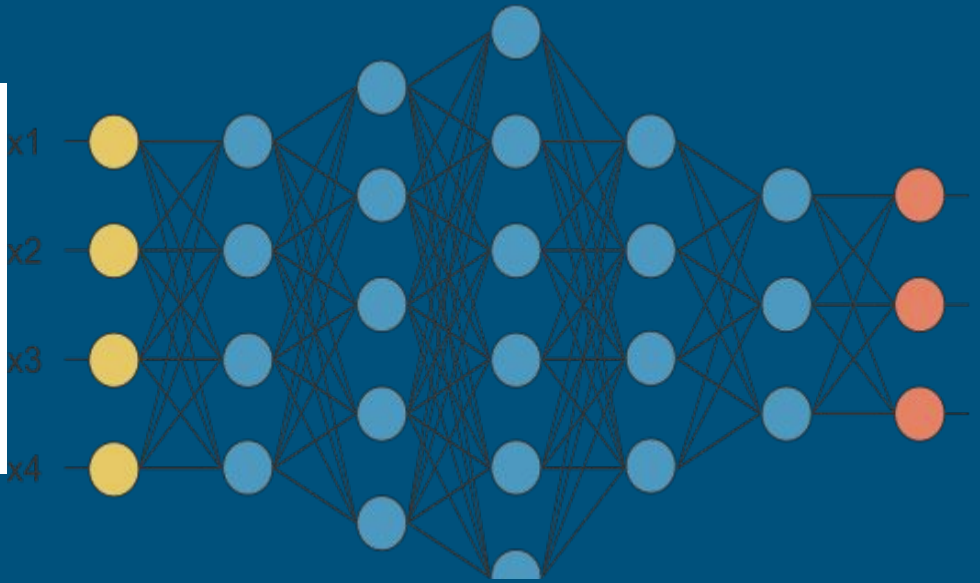
Fielding Independent Pitching (FIP)

Earn Run Average (ERA)

Statcast

Maximum a Postiori (MAP)

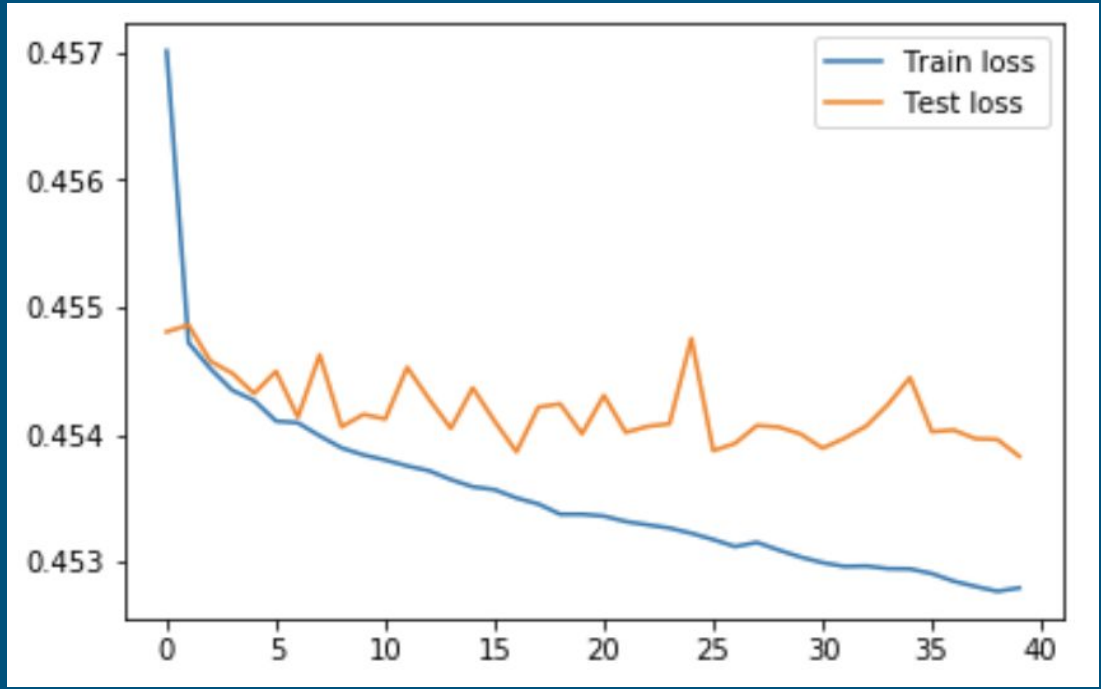
Predicting Success with Neural Networks



Model One: Binary Classification

Features:

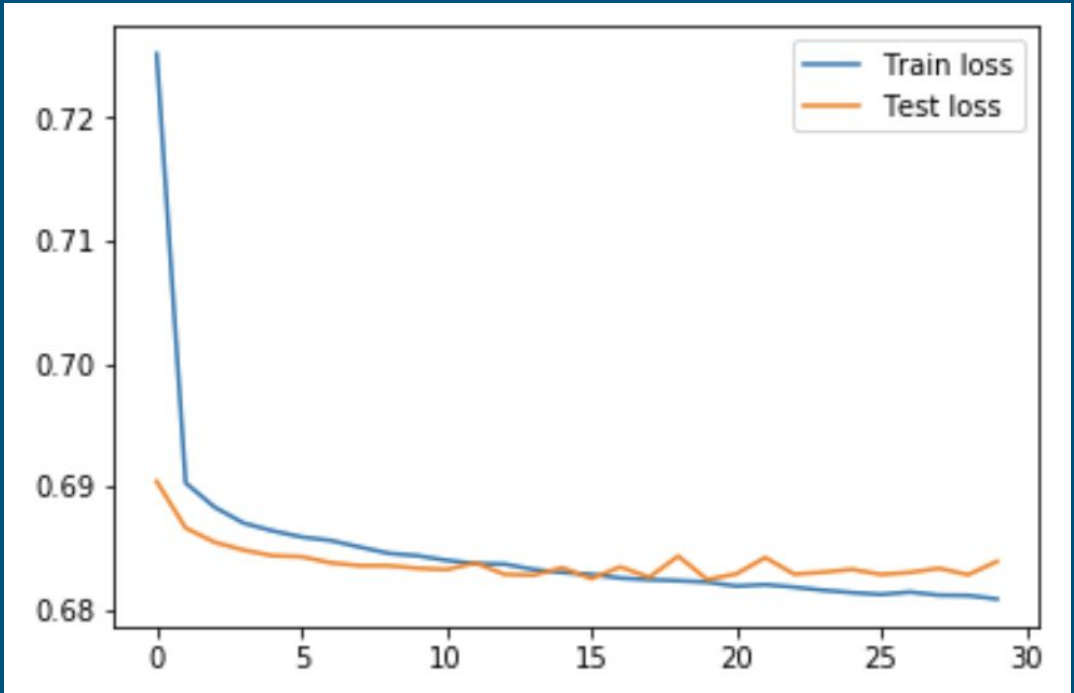
- **Release Speed**
- **Release pos_x**
- **Releasepos_z**
- **Release spin rate**
- **pfx_x**
- **Events**



Model Two: MultiClass Classification

Features:

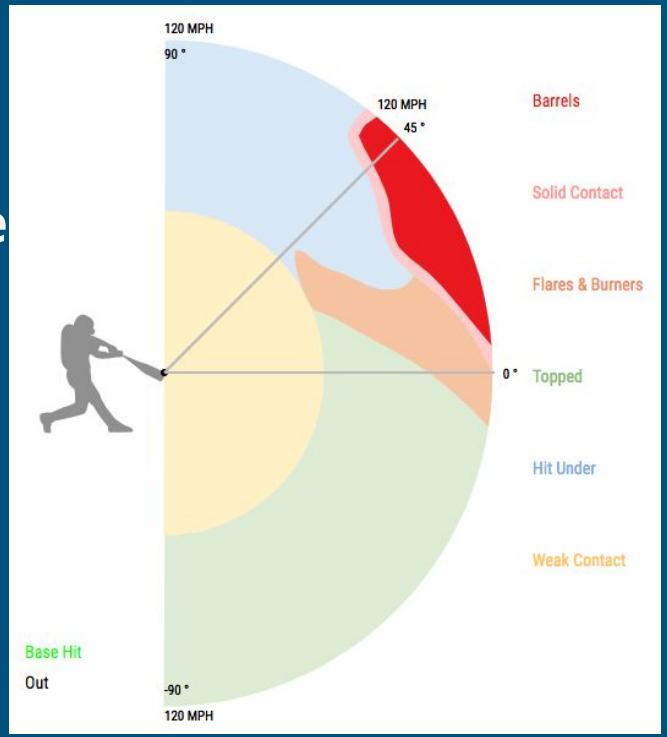
- **Release Speed**
- **Release pos_x**
- **Releasepos_z**
- **Release spin rate**
- **pfx_x**
- **Events**



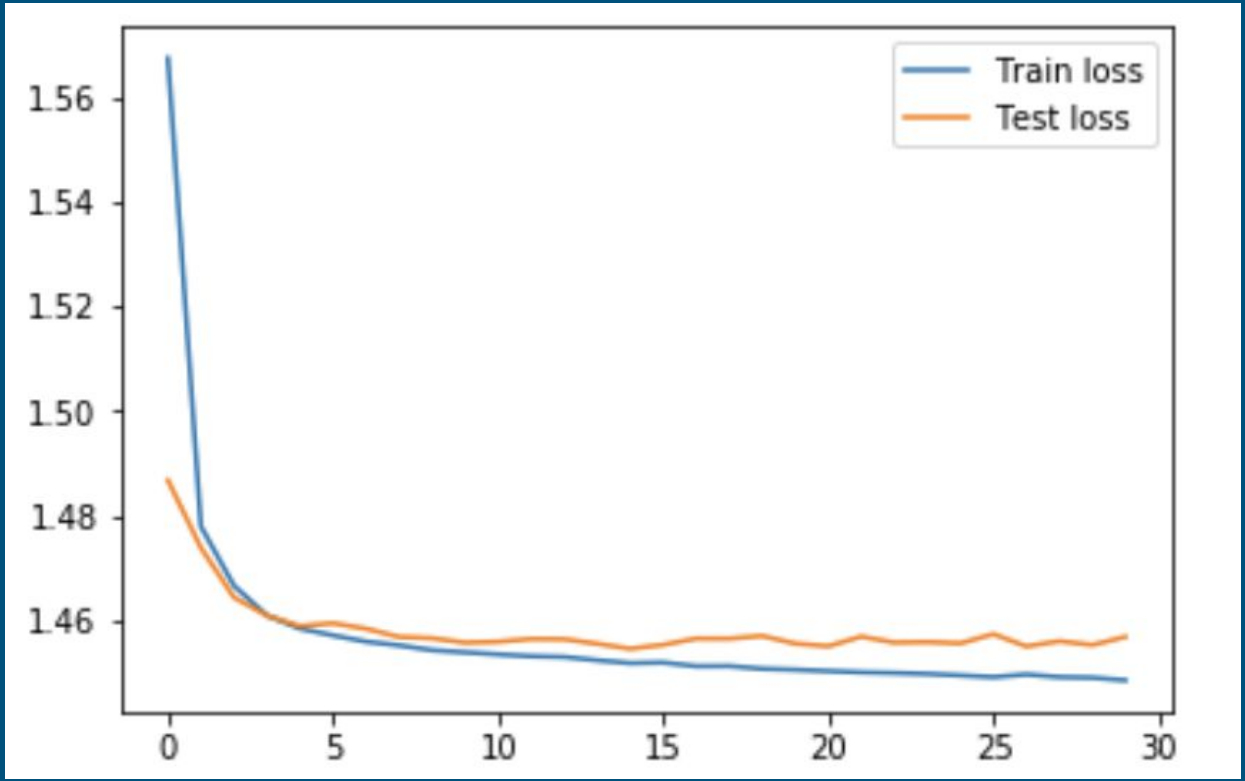
Model Three: MultiClass Classification 2016-18

Features:

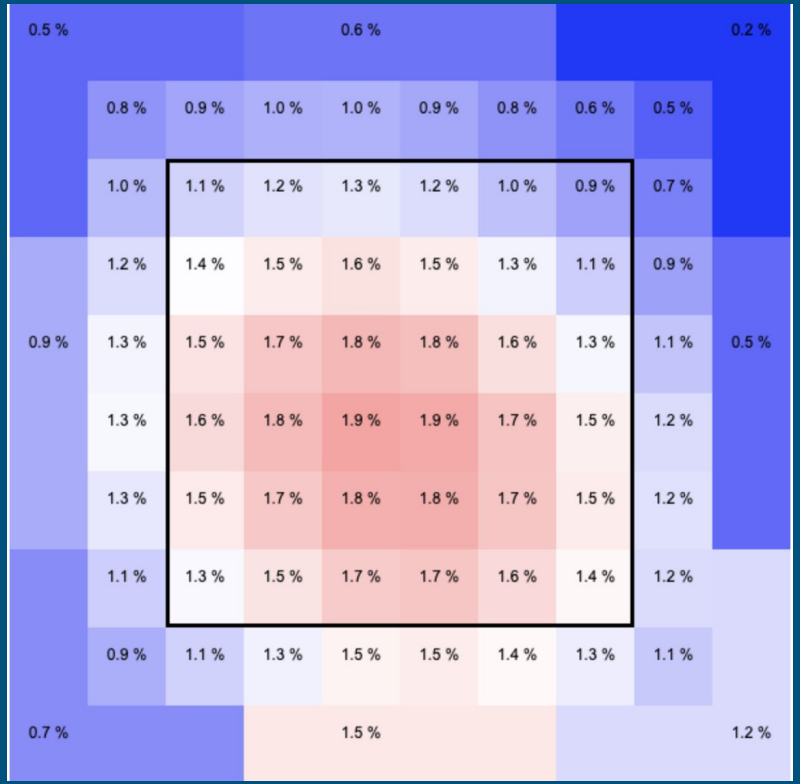
- Plate x
- Plate z
- ax
- ay
- Effective speed
- Release speed
- Release pos_x
- Release_pos_z
- Release spin rate
- pfx_x
- pfx_z
- Launch speed angle



Model 3 Cont...

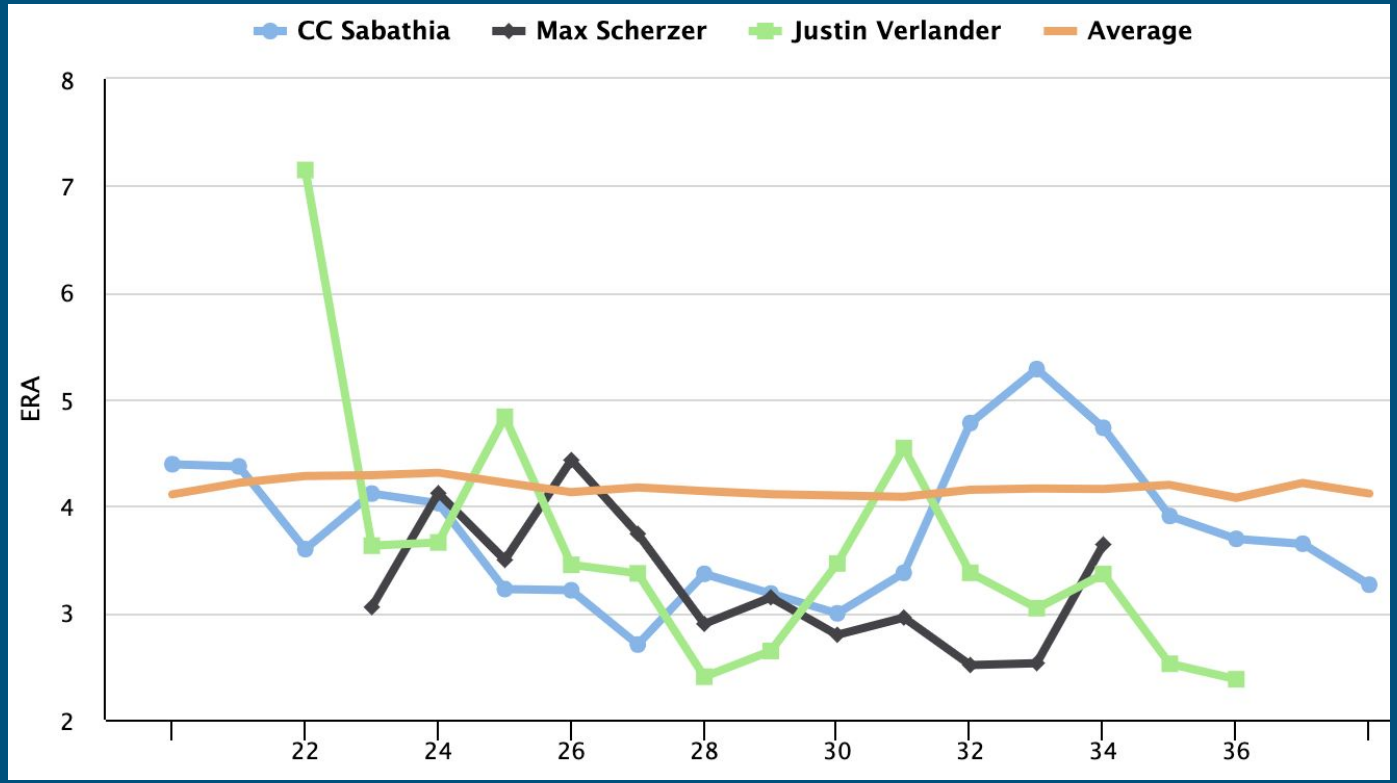


Pitch Heatmap 2016 Season



Goal Two: Use Bayesian Methods to Predict ERAs for the 2019 season

ERA over 14 seasons



Find the MAP for 2019

Name	ER	IP	ERA	MAP
Justin Verlander	60.0	214.0	2.52	2.678808
Justin Verlander	77.0	227.2	3.04	3.052030
Justin Verlander	77.0	206.0	3.36	3.272109
CC Sabathia	78.0	179.2	3.91	3.633982
CC Sabathia	62.0	153.0	3.65	3.431535
CC Sabathia	61.0	148.2	3.69	3.463167
Max Scherzer	62.0	220.2	2.53	2.683323
Max Scherzer	56.0	200.2	2.51	2.682165
Max Scherzer	75.0	228.1	2.96	2.986397

Conclusions

Neural Networks are extremely powerful

Binary Models are more accurate but can be underfit easily with Statcast Data

Gradient Descent was successful with both the all three models.

Pitches that are outside the strike zone are still dangerous pitches to throw if over the plate.

Justin Verlander should expect to have the best year of the three with only a .15 increase to his ERA

Bayesian did not quite predict Scherizer's year