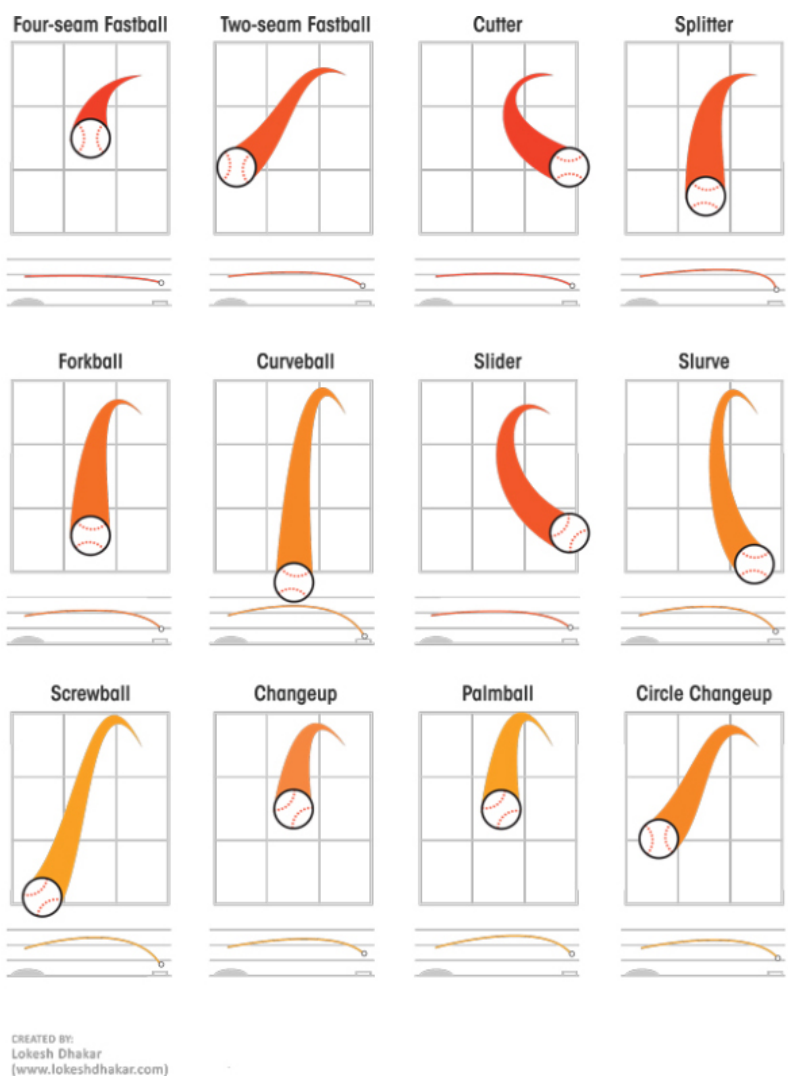
Give 'em the heat, Ricky!
Predicting the next pitch in Major League Baseball

Your Name
Date

Pitch Prediction

In the game of baseball, its hitter versus pitcher. If the hitter knows what pitch is coming next he has a better chance of putting the ball in play. Most starting pitchers use four pitches, fastball, curve ball, slider, changeup, or a specialty pitch such as a sinker, knuckle. These pitches appear to the eye like: http://www.thecompletepitcher.com/images/baseball_pitches.jpg.

The idea of this project will be to determine if pitch count dictates which pitch is thrown.



Previous Research

In March 2012 Ganeshapillai released a study at a sports conference using machine learning to predict situational pitching. He factored in pitch count, runners on base, outs, big lead v small lead, and pitcher handedness v batter handedness. He trained his data on 2008 data, and used 2009 as his untrained data. He also used only pitchers who threw 300 pitches in both season for an even comparison. He used a SVM-Light tool to test his theory, and improved it over a traditional model.

In 2015 Bock, followed up on his research and tested similar factors, this time with a SVM formulation with linear and radial basis function kernels, and a linear logistic model. He got similar answer to Ganeshapillai.

Algorithms

SVM – Light tool

SVM formulation with linear and radial basis function kernels

Evaluation

I hope to be able to repeat the success, if not improve upon it, by Ganeshapallai and Brock.

Data Sources

PitchR/x Cran- R Library (https://cran.r-project.org/web/packages/pitchRx/pitchRx.pdf)
MLB Gameday API
BaseballReference.com

Bock, J. R. (2015). Pitch Sequence Complexity and Long-Term Pitcher Performance. Sports, 3(1), 40-55. Accessed (2/29/2016 http://www.mdpi.com/2075-4663/3/1/40/htm)

References

Ganeshapillai, G.; Guttag, J. Predicting the Next Pitch. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 2 March 2012. Accessed(2/29/16 http://www.theebgar.net/wp-content/uploads/2011/04/Predict-the-next-pitch.pdf)

Fry, M.; Ohlmann, J. Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications.Interfaces 2012, 42, 105–108. [Google Scholar] [CrossRef]

Kalk, J. Pitch Sequence: High Fastball Then Curveball. Hardball Times. Available online: http://bit.ly/1zWrSbb

Glonek, G.; McCullagh, P. Multivariate Logistic Models. J. R. Stat. Soc. Ser. B 1995, 57, 533–546. [Google Scholar]

Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297. [Google Scholar] [CrossRef]

Schölkopf, B.; Smola, A.; Williamson, R.; Bartlett, P. New support vector algorithms. Neural Comput. 2000, 12, 1083–1121. [Google Scholar] [CrossRef]

Rifkin, R.; Klautau, A. In defense of one-vs-all classification. J. Mach. Learn. Res. 2004, 5, 101–141. [Google Scholar]

McCracken, V. Pitching and Defense: How Much Control Do Hurlers Have? Available online: http://bit.ly/1Fel8Zy

Tango, T. Defensive Responsibility Spectrum (DRS): Breaking up the Fielders from the Pitchers. Available online: http://bit.ly/1zFaXFg .

Feder, M.; Merhav, N. Relations between entropy and error probability. IEEE Trans. Inf. Theory 1994, 40, 259–266. [Google Scholar] [CrossRef]

Hughes, G. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 1968, 14, 55–63. [Google Scholar] [CrossRef]

Inoue, A.; Kilian, L. In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? Econom. Rev.2005, 23, 371–402. [Google Scholar] [CrossRef]

Hsu, C.; Lin, C. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. 2002, 13, 415–425. [Google Scholar] [PubMed]

Swartz, M. Game Theory and Baseball, part 2: Introduction to Pitch Selection. Available online: http://bit.ly/1ArqG1I.

Solomonoff, R. A Formal Theory of Inductive Inference (Parts I, II). Inf. Control 1964, 7, 1–22; 224–254. [Google Scholar] [CrossRef]

Kittler, J.; Hatef, M.; Duin, R.; Matas, J. On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 226–239. [Google Scholar] [CrossRef]

Mukherjee. S and Vapnik. V, "Multivariate density estimation: A support vector machine approach". Technical Report, AI Memo 1653, MIT AI Lab

T. Joachims, et al, Making large-Scale SVM Learning Practical, "Advances in Kernel Methods - Support Vector Learning", MIT-Press, 1999.

Robert M. Bell and Yehuda Koren, "Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights". In Proceedings of the 2007 Seventh IEEE International Conference on Data Mining (ICDM '07).