A Sparse Regression Approach For Evaluating and Predicting NHL Results

Nicholas Laffey
Undergraduate Research Assistant
University of Alabama
nblaffey@crimson.ua.edu

Brendan Ames
Assistant Professor, Department of Mathematics
University of Alabama
bpames@ua.edu

December 1, 2016

Abstract

We consider the use of sparse regression techniques in the development of tools for predicting and analyzing NHL team performance. In particular, we develop generalized linear models based on OLS and Poisson regression with elastic net regularization for estimating the number of regular season and playoff wins from a wide variety of regular season team statistics. Our models outperform most previous models in terms of playoff series prediction accuracy. Moreover, the use of the sparsity-inducing elastic net penalty thresholds almost all coefficients in our models to zero, providing increased interpretability; the nonzero coefficients correspond to the statistics that are influencing the model's estimate of team performance.

1 Introduction

Folkloric claims are often made to distinguish truly excellent and championship-caliber sports teams. For example, teams with strong defenses are often favored, i.e., "defense wins championships", as are teams with more experience. However, relatively little evidence is available to establish that such claims are accurate or these effects, if present, are causal rather than correlated with some greater unknown effect. The goal of this work is to partially address this gap in knowledge via the development of sparse regression models for the prediction of regular season and playoff wins for NHL teams. In particular, these models will combine models predicting the number of regulation wins each NHL team earns, with feature selection to provide intuition as to which team statistics are most important to this prediction.

Predicting performance in hockey faces several significant challenges. Most significant is that low scoring rates cause significant randomness and variance in results. Moreover, the continuous nature of play and frequent changes of possession causes difficulty in estimating dominance of puck possession and, hence, of likelihood of winning. In recent years, the use of shot counts as surrogates for puck possession has become extremely popular among hockey analysts. The simplest such is to simply count shots on goal and use shot differential (shots for minus shots against), however Fenwick (shots and missed shots) and Corsi (shots, missed shots, and blocked shots) differential have gained popularity and prominence. These statistics are less scarce than goals and thus provide a more robust estimate of future team performance than goal differential and, by extension, counts of wins and losses; see [Macdonald, 2012c] and the references within. In particular, we can calculate a crude estimate of expected goal differential using shot, Fenwick, or Corsi counts and average shooting percentage rates.

In [Macdonald, 2012c], Macdonald extends this formula for expected goals to include additional offensive statistics (goals, shots, missed shots, blocks shots, Fenwick rating, Corsi rating, zone starts, turnover rates, faceoff rates, hits, shooting percentage). In particular, Macdonald fits a linear model for goals to these predictor variables using ridge regression. The intuition behind such an approach is that the inclusion of these additional variables will provide information about the quality of shots a team or player takes and, hence, a more accurate estimation of the rate at which goals are scored. This approach was subsequently refined and extended in [Arndt and Brefeld, 2016, Macdonald, 2012a, Macdonald, 2012b, Macdonald et al., 2012].

Taking this intuition to its natural extreme, we use a kitchen sink approach to developing regression models for analyzing team performance. Specifically, we train models for estimating the number of regular season and playoff wins using a much wider variety of statistics as predictor variables. Where MacDonald used eleven team statistics, we use fifty three statistics split over multiple scenarios, such as when the team is leading, trailing, one teams lead is within two goals, the team is shorthanded, etc., as well as statistics regarding distribution of salary cap usage and average player ages; we train using 1075 predictor variables in total. The rationale for such an approach is simple: teams adjust strategy depending on the situation. For example, when one team has a significant lead they may tend to play more defensively or conservatively to protect their lead and statistics such as shot attempts and scoring chances will be depressed; this has the illusory effect of the leading team playing poorly, while the opposite is almost certainly true in practice. A full list of team statistics and predictor variables used is available from http://bpames.people.ua.edu/publications.html.

An unfortunate natural consequence of this expansion of predictor variables is the introduction of small sample size issues. Specifically, in order to accurately train 1075 dimensional linear models, we would naively expect to need significantly more than 1075 training observations; however, recorded data for these statistics is limited to the last decade or so worth of NHL seasons (or less than 500 team observations). On the other hand, it is reasonable to expect that many of these statistics are heavily correlated. For example, it is incredibly unlikely for a team to have high Fenwick score while having a low Corsi score, or low shots on goal since these statistics differ only by the addition or deletion of certain types of shot attempts in the count. Moreover, Fenwick differential, for example, is a linear combination of Fenwick attempts for and Fenwick attempts against. This suggests that these high-dimensional observations actually depend on relatively few independent factors. To accurately capture predictive models exploiting this hidden low-dimensionality, we introduce dimension reduction via regularization using sparsity-inducing penalty functions. This has the added effect of providing feature selection to our models. The final models will use only a relatively small number of the predictor variables with nonzero coefficients although which variables is unknown until after calculating the model; those variables with nonzero coefficients in the model will correspond to the team statistics most relevant to predicting team performance and distinguishing between quality of play.

2 Sparse regression models

We use both ordinary least squares (OLS) regression and Poisson regression to train generalized linear models. Specifically, in OLS regression, we assume that the number of wins a team earns, which we treat as a dependent variable, depends linearly (or affinely) on the predictor or independent variables defined by the collected team statistics. Suppose that we have n teams to train our model using $p = 1075$ predictor variables or team statistics. To learn the weighting of predictor variables that minimizes the sum of squared error between the actual number of wins and that predicted by the linear model, we solve the optimization problem

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \; \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2, \tag{2.1}$$

where $\boldsymbol{y} \in \mathbf{R}^n$ is the vector encoding the number of team regulation and overtime wins, i.e., y_i is the number of regulation and overtime wins team i earns, $\boldsymbol{X}$ is the $n \times p$ data matrix whose rows store the predictor variable values for all n teams, and the decision variable $\boldsymbol{\beta}$

yields the optimal weighting of team statistics which minimizes prediction error; here $\|\boldsymbol{z}\|_2$ denotes the standard vector norm on $\mathbf{R}^p$, defined by $\|\boldsymbol{z}\|_2 = \sqrt{z_1^2 + z_2^2 + \cdots + z_p^2}$. We omit shootout wins from the response variable $\boldsymbol{y}$ to eliminate any random variance in win totals introduced by the shootout; we effectively treat a shootout win or loss as a tie. We treat each set of team statistics for each team and year as a separate observation; for example, we have distinct rows of $\boldsymbol{X}$ and entries of $\boldsymbol{y}$ corresponding to the 2012-13 and 2011-12 rosters of the Anaheim Ducks, as well as separate records for the 2012-13 Anaheim Ducks and Boston Bruins.

The number of wins a team may earn is almost certainly not linearly dependent on its predictor variable values, in particular, because teams cannot earn fractional or negative numbers of wins. We train a Poisson regression model to address this nature of the dependent variable. Specifically, in our Poisson regression model we assume that the number of wins follows a Poisson distribution and the logarithm of its expected value can be modeled as a linear combination of predictor variables. To learn the best choice of parameters in this linear function, one maximizes the likelihood of obtaining the desired number of wins given the parameters by solving the optimization problem

$$\max_{\boldsymbol{\theta} \in \mathbf{R}^p} \sum_{i=1}^{n} \log\left(p(y_i; e^{\boldsymbol{\theta}^T \boldsymbol{x}})\right), \tag{2.2}$$

where $p(y_i; e^{\boldsymbol{\theta}^T \boldsymbol{x}})$ is the probability mass function of the Poisson distribution with mean set equal to $e^{\boldsymbol{\theta}^T \boldsymbol{x}}$ for choice of parameters $\boldsymbol{\theta} \in \mathbf{R}^p$; we direct the reader to Chapters 3 and 9 of [Friedman et al., 2001] for more detailed discussion of ordinary least squares regression and Poisson regression respectively.

Since the number of training observations n is much less than the number of unknown parameters p, both generalized linear models provided by the linear regression and Poisson regression are prone to overfitting. That is, it is possible to pathologically choose values of our parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ that make the in-sample error virtually zero, but consequently limit the predictive performance of the generalized linear model on out-of-sample data. To avoid overfitting and to introduce a feature selection element, we add regularization in the form of an elastic net penalty to the optimization problems (2.1) and (2.2):

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\left((1-\alpha)\|\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1\right) \tag{2.3}$$

$$\max_{\boldsymbol{\theta} \in \mathbf{R}^p} \sum_{i=1}^{n} \log\left(p(y_i; e^{\boldsymbol{\theta}^T \boldsymbol{x}}\right) - \lambda\left((1-\alpha)\|\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1;\right) \tag{2.4}$$

here, $\|\boldsymbol{\beta}\|_1$ denotes the ℓ_1-norm on $\mathbf{R}^p$, equal to the sum of the absolute values of the entries of $\boldsymbol{\beta}$, $\|\boldsymbol{\beta}\|_1 = |\beta_1| + |\beta_2| + \cdots + |\beta_p|$, which acts as a surrogate for the number of nonzero entries of $\boldsymbol{\beta}$, $\lambda > 0$ is a regularization parameter representing the strength of the penalty term, and $0 < \alpha < 1$ controls which of the two penalty terms is emphasized in the elastic net. Solving (2.3) and (2.4) essentially seeks a vector of parameters $\boldsymbol{\beta}$ or $\boldsymbol{\theta}$ that compromises between minimizing error or maximizing likelihood, respectively, and minimizing the number of nonzero parameters, as controlled by the elastic net penalty function $(1-\alpha)\|\boldsymbol{\beta}\|_2^2+\alpha\|\boldsymbol{\beta}\|_1$; the parameter λ tunes between these two competing objectives. For sufficiently large choice of λ and α, the ℓ_1-norm penalty thresholds almost all entries of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to zero, effectively limiting the search for parameters to relatively few predictor variables, which significantly reduces overfitting; the few remaining predictor variables with nonzero coefficients are those features and statistics which are used in the ultimate generalized linear models. A comprehensive discussion regarding the use of sparsity-inducing penalties in regression is beyond the scope of this article; we direct the reader to the recent monograph [Hastie et al., 2015] for a more detailed discussion of this approach and the use of sparsity in statistical learning in general.

3 Predictions

3.1 Training two new models

We collected detailed NHL team statistics for the seasons from 2007-08 to 2014-15. We divide the statistics from these seasons into training data sets which are used to generate generalized linear models using elastic net regularized ordinary least squares and Poisson regression, and test sets used to validate our models as follows:

1. *Regular season.* Using sparse OLS and Poisson regression we train generalized linear models to fit the number of regulation and overtimes wins each team earns to detailed regular season statistics. We then use the model to predict the number of wins each team in the test sets

2. *Playoffs.* We also developed models for predicting playoff performance using regular season team statistics. Specifically, we fit sparse OLS and Poisson regression generalized linear models to the set of regular season training data restricted to teams which made the playoffs, with number of playoff wins used as a response variable. We then predict the number of playoff wins for each out-of-sample playoff team in the test set by substituting their regular season statistics into our models.

All models were created using the `glmnet` package [Friedman et al., 2009] in `R`. We choose values of the regularization parameters α and λ in the elastic net using 10-fold cross validation.

	West				**East**			
	Poisson		Actual		Poisson		Actual	
1.	*VAN*	47.18	SJ	46	**WSH**	56.84	WSH	49
2.	**CHI**	47.06	CHI	46	*BUF*	41.21	NJ	45
3.	*SJ*	46.73	VAN	45	*NJ*	40.17	BUF	40
4.	*LA*	38.33	ARI	45	**PIT**	37.58	PIT	37
5.	COL	37.90	DET	40	*PHI*	36.53	OTT	42
6.	*DET*	36.19	LA	36	*NYR*	35.45	BOS	37
7.	ARI	35.55	NSH	41	*BOS*	34.25	PHI	38
8.	*STL*	34.16	COL	37	**MTL**	32.59	MTL	36
9.	*CGY*	33.48	STL	36	OTT	31.84	NYR	29
10.	*DAL*	32.81	CGY	31	**WPG**	31.02	WPG	30
11.	NSH	32.60	ANA	35	**CAR**	30.50	CAR	30
12.	*ANA*	32.43	DAL	32	*FLA*	28.79	TB	28
13.	**MIN**	31.45	MIN	35	**NYI**	27.88	NYI	30
14.	**CBJ**	28.41	CBJ	27	*TB*	27.67	FLA	28
15.	**EDM**	24.35	EDM	25	**TOR**	25.12	TOR	25

Table 3.1: Standings predicted by the sparse Poisson model and actual standings for the 2009-10 season. Predicted position of bold teams matches actual position in standings. Teams in italics are within 2 positions of their actual position.

Each predictor variable was normalized and centered to have mean 0 and variance 1 to avoid any distortion in the models due to differing scales of statistics. `R` code and detailed results from our experiments can be found at http://bpames.people.ua.edu/publications.html.

3.2 Regular season results

We train our regular season models using detailed team statistics from the 2011-12, 2013-14, and 2014-15 seasons. We then use our models to predict the number of regulation and overtime wins of each team for the 2009-10 NHL season. Table 3.2 presents the regular season standings for the 2009-10 NHL season as predicted by our sparse Poisson model.

The predicted standings are remarkably faithful to the actual standings. The sparse Poisson model accurately predicts fourteen of the sixteen playoff positions. Moreover, all but four ofthe teams is ranked within two positions of their actual standing and fourteen of thirty are ranked by the Poisson model at the same position as their actual standing. Some of these discrepancies can be explained by the NHL's use of a point system (2 for a win, 1 for a loss in overtime or shootout, 0 for a regulation loss) rather than strict wins and losses for ranking teams. For example, the Philadelphia Flyers had the fifth most regulation and overtime wins in the Eastern Conference yet were ranked only seventh by points; our ranking places Philadelphia fifth by Poisson wins, which agrees with their ranking by regulation and overtime wins. Moreover, the difference between the predicted number of wins and the actual number of wins is 0.90 for the Colorado Avalanche, yet Colorado ranked three positions higher in the Poisson rankings.

There were three significant outliers in our rankings, the Arizona (then Phoenix) Coyotes, Nashville Predators, and the Ottawa Senators. All three were ranked significantly lower in the Poisson standings than they were in the actual standings. This suggests that these teams may have obtained significantly better records than they deserved due to their play. For example, in our model Ottawa moves out of a comfortable playoff position, which is supported by their poor goal differential of -13. Both Arizona and Nashville had roughly 10 more wins than predicted by the Poisson model. On the other hand, our model suggests that the New York Rangers and St. Louis Blues may have deserved playoff positions due to their play despite missing the playoffs; in particular, the New York Rangers underperformed their predicted win total by over six wins, removing them from a playoff position.

3.3 Playoff predictions

We conducted a second analysis where we compare the predictive power of our generalized linear models with several popular existing statistics. We train our models using team statistics from the years 2009-10, 2010-11, 2011-12, and 2013-14. Specifically, we fit the number of playoff wins for each team that qualified for the playoffs to their regular season team statistics; In addition to sparse OLS and sparse Poisson regression models, we also train OLS and Poisson regression models with ridge regression, i.e., instead of the elastic net penalty, we use the Euclidean norm $\|\boldsymbol{\beta}\|_2$ and $\|\boldsymbol{\theta}\|_2$ in (2.3) and (2.4). This new approach applies regularization to address the underdetermined nature of the fitting problem, but without the additional feature selection provided by the ℓ_1-norm. This yields models that are less interpretable, since they use all predictor variables, but with improved classification performance.

We use these generalized linear models to predict t future playoff performance using each team's regular season performance. We use data from the 2007-08, 2008-09, 2012-13, and 2015-16 seasons to test our trained models' predictive ability. For each playoff team from each of these seasons, we calculate its expected number of playoff wins as given by the each

Method	Success Rate
Poisson Ridge	0.6833
OLS Ridge	0.6833
Goal Differential	0.6833
Fenwick	0.6833
Expected Goals	0.6667
Corsi	0.6167
Poisson Elastic Net	0.6000
Wins	0.6000
OLS Elastic Net	0.5778
PDO	0.4833
Team Salary	0.4667

Table 3.2: Prediction accuracy of statistics for playoff series. Success rate is the fraction of playoff series where the team with the larger value of predictor wins.

generalized linear model. We then compared the expected wins for each team in each playoff series and declare the predicted winner to be the team with the most expected wins. We repeated the process for a number of other predictive statistics (Goal differential, Fenwick, MacDonald's Expected Goals metric, Corsi, regular season wins, PDO (save percentage plus shooting percentage), and percentage of salary cap used).

Table 3.3 summarizes the results of this trials. Note that the OLS and Poisson models constructed using ridge regression yield the best out-of-sample predictive performance (matched by goal differential and Fenwick scores), while their counterparts calculated using elastic net regression perform slightly worse. This suggests that there is a cost in predictive ability for using more interpretive models. Moreover, the ridge regression models with expanded predictor variables outperform the Expected Goal approach; this supports the intuition that using more information would lead to improved prediction performance.

3.4 Interpreting the model coefficients

An advantage of employing elastic net regression is the increased interpretability of the generalized linear models. In this case, our linear models use between 20 statistics, in the Poisson regression model for predicting playoff performance, and 49 statistics, in the OLS model for estimating regular season wins, out of 1075 total statistics. Using such a small number of statistics provides intuition into what properties are distinguishing team

Season	Reg	Off	Def	Both	Hits	Pen	ZS	FO	Sal	Card
Regular	OLS	0.3673	0.1633	0.1429	0.0204	0.1020	0.1224	0.0612	0.0204	49
Regular	Pois	0.3684	0.1842	0.1579	0.0263	0.0526	0.1316	0.0526	0.0263	38
Playoffs	OLS	0.3056	0.2500	0.2222	0.0000	0.0556	0.0556	0.0833	0.0278	36
Playoffs	Pois	0.3500	0.2500	0.2000	0.0000	0.0500	0.0500	0.0500	0.0500	20

Table 3.3: Distribution of nonzero coefficients in generalized linear models. Coefficients are grouped by type: Offensive (Off), Defensive (Def), combinations of offensive and defensive statistics (Both), body checking (Hits), penalty counts (Pen), zone starts (ZS), face-off proficiency (FO), salary and age (Sal). The column "Card" provides the total number of nonzero coefficients in each linear model. Each entry provides the fraction of nonzero coefficients in each model.

performance. Table 3.4 summarizes the distribution of these statistics.

One observation that can be made from the model coefficients is that regular season defensive performance is significantly more influential in predicting playoff performance than in the regular season; statistics reflecting a combination of both offensive and defensive ability also exhibit increased influence in models fit to playoff performance. This is most acutely noticeable in the collection of model coefficients contained in Table 3.4, which is obtained by retraining our sparse linear models with the added restriction that between 8 and 15 predictor variables be employed by the model and our training data restricted to cumulative statistics over all scenarios. Inspecting the coefficients in the elastic net Poisson regression models indicates that standard offensive statistics, such as shooting percentage and goals for have little influence over the models' prediction of playoff performance, while heavily influencing the estimation of regular season wins; in fact, high scoring rates negatively impact the predicted playoff performance in the Poisson model. Similarly, the total number of scoring chances allowed and obtained negatively influences predicted playoff performance. This suggests that successively playing conservatively, i.e., limiting offensive chances against at the expense of scoring, is more important in the playoffs than in the regular season, lending credence to the cliche "defense wins championships". Surprisingly, goal differential is not used in the elastic net Poisson model, despite regular season goal differential having similar predictive success to the ridge regression Poisson model.

Metric	Regular	Playoff	Type
Shooting %	0.027761888	0	Offensive
Goals For	0.006295492	-0.009847594	Offensive
Goals For %	0.107205337	0	Offensive
Power Play Goals Allowed	0	-0.005191436	Defensive
Penalty Killing %	0	0.02301938	Defensive
Save %	0	0.015758699	Defensive
Scoring Chances Against	0	-0.008546852	Defensive
Goals Allowed	0	-0.036184537	Defensive
PDO	0.006788737	0	Combined
Goal Differential	0.055360021	0	Combined
High Danger Scoring Change +/-	0.012927707	0	Combined
Scoring Chances For and Against	0	-0.007569431	Combined

Table 3.4: Coefficients of select statistics in elastic net Poisson regression models.

4 Conclusions and Future Research

We have developed new models for predicting NHL team performance based upon penalized OLS and Poisson regression. Empirical evidence suggests that these methods are competitive with the current state of the art, while the increased interpretability of our models provides evidence regarding what factors are influencing the predictions and, perhaps, what distinguishes successful teams in both the regular season and playoffs. Although our current work focuses on NHL statistics, there does not seem to be an obvious barrier to extending this approach to other hockey leagues, except perhaps a lack of readily available raw statistics, or other sports leagues.

A natural extension of our approach would be to develop similar models for analyzing the performance of individual players. This could take the form of an expected goals and adjusted plus-minus metric based on elastic net regression similar to the models proposed by MacDonald in [Macdonald, 2012c]. On the other hand, fitting minor league or early career performance to total career performance, as measured by career games played, total goals scored, etc., could provide a valuable scouting tool in drafting and developing young players. Finally, fitting team performance, as measured by goals, wins, points, etc., using elastic net regression to individual player statistics could provide a means of attributing team success to individual players and a useful tool for lineup construction.

Another improvement would be to use boot strapping to obtain more robust models. For example, we could randomly sample subsets of our training seasons, fit our models, and repeat many times; we'd then average the coefficients to obtain the final models. This approach would yield models that are less sensitive to variation between seasons, but at a possible cost of reduced interpretability. On the other hand, limiting our training and testing sets to pairs of consecutive seasons and fitting sparse regression models might reveal how influential predictive statistics evolve over time and, consequently, how properties of successful teams are effected as rules and strategy vary from season to season.

References

[Arndt and Brefeld, 2016] Arndt, C. and Brefeld, U. (2016). Predicting the future performance of soccer players. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(5):373–382.

[Friedman et al., 2001] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

[Friedman et al., 2009] Friedman, J., Hastie, T., and Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1.

[Hastie et al., 2015] Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC Press.

[Macdonald, 2012a] Macdonald, B. (2012a). Adjusted plus-minus for nhl players using ridge regression. *URL http://arxiv. org/abs/1201.0317 v1.*

[Macdonald, 2012b] Macdonald, B. (2012b). Adjusted plus-minus for nhl players using ridge regression with goals, shots, fenwick, and corsi. *Journal of Quantitative Analysis in Sports*, 8(3):1–24.

[Macdonald, 2012c] Macdonald, B. (2012c). An expected goals model for evaluating nhl teams and players. In *Proceedings of the 2012 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference. com.*

[Macdonald et al., 2012] Macdonald, B., Lennon, C., and Sturdivant, R. (2012). Evaluating nhl goalies, skaters, and teams using weighted shots. *arXiv preprint arXiv:1205.1746.*