# Ranking Sports Teams

Jason D. M. Rennie
jrennie@gmail.com

February 2, 2005

**Abstract**

Compared to collegiate sports, professional sports teams' schedules are balanced and comprehensive. Hence, in professional sports, it is usually reasonable to rank teams by their winning percentage. Not true in collegiate sports where two teams with identical records may have very different skill levels. The problem is that there are well over 300 Division-I teams, yet each plays only ten or twenty unique teams. Most teams only play a few teams outside their conference, yet there are over 30 conferences. The #1 team in the Metro Atlantic Athletic Conference may not stand a chance against the last place Big 12 team. In short, a team's winning percentage is not enough to determine how good it is. We need something that accounts for who their opponents were. In this paper, we associate each team with a weight that represents how good the team is. We postulate a model wherein the probability that one team beats another is the difference in their weights, passed through the logit function. We learn weights for teams that maximize the likelihood of the known outcomes. The learned weights can then be used to determine team rankings (by sorting teams according to their weights); the weights can also be used to predict the winner of a future contest and the estimated probability of one team beating the other.

We assume that there is a set of data describing past contests, $X \in \{+1, 0, -1\}^{n \times l}$. That is $X$ is a matrix of $n$ rows (one for each contest) and $l$ columns (one for each team), where entries take on one of three values: positive, negative or zero. $X_{ij} = +1$ if the home team in contest $i$ was team $j$; $X_{ij} = -1$ if the away team in contest $i$ was team $j$; $X_{ij} = 0$ if team $j$ did not play in contest $i$. In short, each row of $X$ has two non-zero entries indicating the home $(+1)$ and away $(-1)$ teams. We also assume that we have a vector, $\vec{y} \in \{+1, -1\}^n$, telling us who won each game. $y_i = +1$ if the home team won contest $i$; $y_i = -1$ if the away team won contest $i$. Note that this model does use home/away information, so if a game is played at a neutral site, the home/away designation can be made arbitrarily. We associate a weight, $\{w_1, \ldots, w_l\}$, with each team and learn weights so as to maximize the likelihood of our model:

$$P(\vec{y}|X,\vec{w}) = \frac{1}{1+\exp(-\vec{y} * (X\vec{w}))} = \prod_{i=1}^{n} \left(1+\exp\left(-y_i \sum_{j=1}^{l} X_{ij} w_j\right)\right)^{-1} \quad (1)$$

Note that this is simply a Logistic Regression model [1]. Since the model is convex [2] we can find the globally optimal weight vector using any gradient-descent-type algorithm that is guaranteed to converge (such as Conjugate Gradients [3].

The learned weight vector can be used to determine a ranking for the teams: simply sort the weights; the team with the largest weight is #1 (according to this model). The weights can also be used to predict the outcome of a future contest: $(1 + \exp(w_j - w_i))^{-1}$ is the probability that team $i$ will beat team $j$.

This model will try to give an infinite weight to any team that has never won ($-\infty$) or never lost ($+\infty$). To avoid this (and to improve generalization to unseen data), we can penalize large weights. We add an L2-penalty to the negative-log-likelihood and learn weights by minimizing this quantity:

$$J = -\sum_{i=1}^{n} \log\left(1+\exp\left(-y_i \sum_{j=1}^{l} X_{ij} w_j\right)\right) + \frac{\lambda}{2}\sum_{j=1}^{l} w_l^2. \quad (2)$$

$\lambda$ is the regularization parameter; it controls the trade-off between the goals of data fitness and small weights. A value of $\lambda = 1$ is not unreasonable.

## References


[1] J. D. M. Rennie. Logistic regression. http://people.csail.mit.edu/~jrennie/writing, April 2003.

[2] J. D. M. Rennie. Regularized logistic regression is strictly convex. http://people.csail.mit.edu/~jrennie/writing, January 2005.

[3] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. http://www.cs.cmu.edu/~jrs/jrspapers.html, 1994.