5 Ranking: Massey’s Original Idea

The following is the idea for ranking put forward in Massey’s undergraduate thesis. (He has more advanced methods now.) By \textit{ranking} we mean a numerical value. From this one can produce ordinal values, if desired.

The method is based on the idea that in a perfect world the teams have numerical ratings such that, in every game, the resultant net score is the difference between the ratings of the teams. That is, there are ratings \( r_i \) such that the margin \( g_{ij} \) of team \( i \) over team \( j \) is given by

\[ g_{ij} = r_i - r_j \]

So the task is, given the actual margins \( g_{ij} \), to determine the \( r_i \). There are of course many more equations than variables; so this set of linear equations almost surely has no solution. Instead, think of the \( g_{ij} \) as a statistic. Then the goal is to choose the vector \( \vec{r} \) to minimize the errors \( g_{ij} - (r_i - r_j) \). The standard approach is to minimize the sum of the squares of the errors. This can be justified by theorems, but more important, it can be computed.

Say there are \( n \) teams and they play \( m \) games. Let \( X \) be the \( m \times n \) game matrix. That is, matrix \( X \) has a row for each game: The row has a 1 in the column corresponding to the winner, a \(-1\) in the column corresponding to the loser, and 0’s elsewhere. Let \( \vec{g} \) be the vector of net margin for each game. Let \( \vec{r} \) be the unknown vector of rankings. Then we have ideally:

\[ \vec{g} = X \vec{r} \]

We want now to minimize the sum of the squares of the entries of \( \vec{g} - X \vec{r} \) (its \( \ell_2 \)-norm). This sum is given by the product \( (\vec{g} - X \vec{r})^t (\vec{g} - X \vec{r}) \). Now, by (matrix) calculus, we can differentiate with respect to \( \vec{r} \) to get

\[ -2(\vec{g} - X \vec{r})^t X = 0 \]

Now, define matrix \( S = X^t X \) and vector \( \vec{N} = X^t g \). Then when the dust settles, we need to solve

\[ S \vec{r} = \vec{N}. \]
Interlude: the matrix $S$ is an $n \times n$ matrix. Each diagonal entry is the number of games that team played; off the diagonal the value is $-1$ if the corresponding teams played, and 0 otherwise. (Why?) Also, the vector $\vec{N}$ is the net score for each team: points-for minus points-against. In particular, we can compute $S$ and $\vec{N}$ directly from the data, and need never construct $X$.

Now, we would be done if we could write $\vec{r} = S^{-1}\vec{N}$. However, there are two problems with this. First, calculating $S^{-1}$ is computationally expensive. But more importantly, $S$ is guaranteed not to have an inverse. There are two ways to see this. The first is that each row of $S$ sums to 0, and thus the all-1 vector is in its null-space. The second is that the solution for $\vec{r}$ is not unique: one can add the same constant to all entries and still have a solution.

So the solution is to discard one row of $S$ (since it is redundant) and add a row that normalizes the rankings. For example, add the equation that the rankings sum to 0. Then go ahead and solve, using some software.

Example: For the 2013 ACC football games, the resulting ratings are

- FloridaSt 35.44
- Clemson 16.90
- GeorgiaTech 7.07
- NorthCarolina 2.88
- VirginiaTech 2.73
- MiamiFL 0.73
- BostonCollege 0.20
- Duke 0.05
- Pittsburgh -3.54
- Syracuse -10.15
- WakeForest -10.67
- Maryland -11.47
- NCState -14.10
- Virginia -16.08
- WakeForest -10.67
- Maryland -11.47
- NCState -14.10
- Virginia -16.08

Mathematical aside: one can justify these steps. For example, one can haul out multivariate calculus (remember Jacobians) and show that $\vec{r}$ is indeed a local minimum. One can show that the adjusted matrix has full rank, so that there is a solution, and so on.

Extensions/Comments. It might be reasonable to use a concave function of $g_{ij}$ (such as $\sqrt{g_{ij}}$) to reduce the impact of large results.

References
