Model 1: Probabilistic Matrix Factorization

This model assumes the number of points scored by team i against team j s_{ij} is Normally distributed with mean $U_i V_j'$ and variance σ_ε^2, where U_i is team i's $m \times 1$ latent offense vector and V_j is team j's $m \times 1$ latent defense vector. We may include the following priors for the latent components and the error variances

$$\begin{aligned} U_i | \mu_u, \Sigma_u &\sim N(\mu_u, \Sigma_u) \\ V_j | \mu_v, \Sigma_v &\sim N(\mu_v, \Sigma_v) \\ \sigma_\varepsilon^2 &\sim IG(\alpha, \beta). \end{aligned}$$

These priors shrink our estimates of the latent factors and provide a baseline for teams for which we have little data. If we have priors, we have the option to include the hyperpriors

$$\begin{aligned} \mu_u | m_u, S_u &\sim N(m_u, S_u) \\ \mu_v | m_v, S_v &\sim N(m_v, S_v) \\ \Sigma_u | d_u, \Psi_u &\sim IW(d_u, \Psi_u) \\ \Sigma_v | d_v, \Psi_v &\sim IW(d_v, \Psi_v). \end{aligned}$$

Hyperpriors build flexibility into the parameters of the prior, so that we don't shrink to completely unrealistic values for the latent factors.

Because this model is somewhat computationally intensive, we only use game data for the 68 teams in the tournament and the play-in. This means that the appropriate choice of priors is essential to the predictive success of the model. Perhaps future implementations can accommodate the full set of games played, in order to limit the informativeness of the priors.

Making Predictions

Conditional on U_i, U_j, V_i, V_j, and σ_ε^2, $s_{ij} - s_{ji}$ is also Normally distributed with mean $U_i V_j' - U_j V_i'$ and variance $2\sigma_\varepsilon^2$. This means i's probability of beating j, conditional on the model parameters is:

$$P(i \text{ beats } j | U_i, U_j, V_i, V_j, \sigma_\varepsilon^2) = \Phi\left(\frac{U_j V_i' - U_i V_j'}{\sqrt{2\sigma_\varepsilon^2}}\right).$$

Denote the set of prior parameters and data by Θ and Y, respectively. Our MCMC procedure yields draws from $P(U_i, U_j, V_i, V_j, \sigma^2_\varepsilon|\Theta, Y)$, which allows us to compute

$$P(i \text{ beats } j|\Theta, Y) = \int P(i \text{ beats } j|U_i, U_j, V_i, V_j, \sigma^2_\varepsilon)\, P(U_i, U_j, V_i, V_j, \sigma^2_\varepsilon|\Theta, Y)\, dU_i\, dU_j\, dV_i dV_j\, d\sigma^2_\varepsilon.$$

If we have hyperpriors for the parameters of our priors, then we have draws from $P(\Theta|\Theta_h, Y)$, where Θ_h is the set of parameters of the hyperpriors, we compute

$$P(i \text{ beats } j|\Theta_h, Y) = \int P(i \text{ beats } j|\Theta, Y) P(\Theta|\Theta_h, Y) d\Theta.$$

Then our predictions are made on the basis of $P(i \text{ beats } j|\Theta_h, Y)$ or $P(i \text{ beats } j|\Theta, Y)$. In each matchup, we choose the team that has a $> 50\%$ chance of winning.

Model 2: Augmented Elo

Just the Basics

Take a model where x_i and x_j are the latent scores of two teams and the probability that team i beats team j is

$$P(i \text{ beats } j|x_i, x_j) = \left(1 + \exp(x_j - x_i)\right)^{-1}.$$

Define the winner of game t as w_t and the loser of game t as l_t. Then, the likelihood of an outcome in game t is

$$P(w_t = i, l_t = j|\mathbf{x}) = \left(1 + \exp\left(\sum_{j=1}^{n} \mathbf{1}(l_t = j)x_j - \sum_{i=1}^{n} \mathbf{1}(w_t = i)x_i\right)\right)^{-1}$$

and the log-likelihood of every game is

$$\log P(\mathbf{w}, \mathbf{l}|\mathbf{x}) = -\sum_{t=1}^{T} \log\left(1 + \exp\left(\sum_{j=1}^{n} \mathbf{1}(l_t = j)x_j - \sum_{i=1}^{n} \mathbf{1}(w_t = i)x_i\right)\right).$$

Because only the differences between xs are identified, we introduce a prior on the xs to normalize their levels and cut down on implausibly decisive forecasts.

Including Margin-of-Victory

If desired, we can work the margin-of-victory into the analysis to incorporate additional information about $\mathbf{x}$. Define $\log m_t$ as the log of the margin-of-victory in game t. We model the conditional expectation for the margin of victory as follows

$$\log m_t = \alpha + \beta \left(\sum_{i=1}^{n} \mathbf{1}(w_t = i) x_i - \sum_{j=1}^{n} \mathbf{1}(l_t = j) x_j \right) + \varepsilon_t$$
$$\varepsilon_t \sim N(0, \sigma^2).$$

Define the expected margin-of-victory and the error in game t as

$$\mu_t(\mathbf{x}, \alpha, \beta) = \alpha + \beta \left(\sum_{i=1}^{n} \mathbf{1}(w_t = i) x_i - \sum_{j=1}^{n} \mathbf{1}(l_t = j) x_j \right)$$
$$e_t(\mathbf{x}, \alpha, \beta) = \log m_t - \mu_t(\mathbf{x}, \alpha, \beta)$$

The log-likelihood of all the margins of victory is

$$\log Q(\mathbf{m}|\mathbf{w}, \mathbf{l}, \mathbf{x}, \alpha, \beta, \sigma) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t(\mathbf{x}, \alpha, \beta)^2.$$

It's possible to introduce informative priors on α, β, and σ^2, though we don't do that in this case. We do, however, maintain the prior on the xs for the same reason we used it in the model without data on the margin of victory.

Efficient likelihood and gradient computation, makes both finding the mode and simulating from the posterior very cheap. As a result, this model, in both variations, uses every single game played during the regular season, unlike the PMF models.

Making Predictions

Denote the set of prior parameters by Θ. Our MCMC procedure yields draws from $P(\mathbf{x}|\mathbf{w}, \mathbf{l}, \Theta)$ and $P(\mathbf{x}, \alpha, \beta, \sigma^2|\mathbf{w}, \mathbf{l}, \mathbf{m}, \Theta)$ for the models with and without margin-of-victory data, respectively. These draws let us compute the integrals

$$P(i \text{ beats } j|\mathbf{w}, \mathbf{l}, \Theta) = \int (1 + \exp(x_j - x_i))^{-1} P(\mathbf{x}|\mathbf{w}, \mathbf{l}, \Theta) d\mathbf{x}$$
$$P(i \text{ beats } j|\mathbf{w}, \mathbf{l}, \mathbf{m}, \Theta) = \int (1 + \exp(x_j - x_i))^{-1} P(\mathbf{x}, \alpha, \beta, \sigma^2|\mathbf{w}, \mathbf{l}, \mathbf{m}, \Theta) d\mathbf{x}\, d\alpha\, d\beta\, d\sigma^2.$$

Then our predictions are made on the basis of $P(i \text{ beats } j|\mathbf{w}, \mathbf{l}, \mathbf{m}, \Theta)$ or $P(i \text{ beats } j|\mathbf{w}, \mathbf{l}\Theta)$. In each matchup, we choose the team that has a $> 50\%$ chance of winning.