Modeling Protein Interaction Experiments with a Trace Norm Prior

Jason D. M. Rennie*
jrennie@gmail.com

December 20, 2006

1 Introduction

One interest in Biology is to determine whether or not a certain pair of proteins will interact given the opportunity. Experiments are conducted which provide partial information about interaction. Typically, each the result of each experiment can be represented as simply a binary indicator function over (a set of) protein pairs. A negative result indicates no interaction. A positive result may be due to one of two things: (1) interaction between the two proteins, or (2) "self-interaction", where one protein is able to produce signs of interaction without help from the second protein. The goal is to uncover true protein-protein interactions (PPIs). So, positive results due to self-interaction are considered noise.

Sontag et al. (2007) proposed a model for capturing this self-interaction noise. They introduced a variable for each protein indicating the chance that self-interaction will yield a positive result. This extra variable per protein allowed the model to "explain-away" cases where the positive result was likely caused by self-interaction and not a true PPI. But, this is not the only source of noise that may lead to incorrect results. A true PPI may not lead to a level of expression that is detected in the experiment. Or, if the experiment is overly sensitive, the experiment may falsely indicate PPI when there is neither self-interaction nor a true PPI. To better handle this, we propose to model how likely a pair of proteins are to interact.

In section 2, we discuss two models of likelihood of the experimental data. Likelihood is the way in which we determine how well our model "fits" the experimental data. In section 3, we discuss the prior we use for the parameters of our likelihood. The prior serves as a regularizer which encourages a compact representation of the data in order to capture relevant aspects of the data and and ignore noise, the idea being that relatively few factors influence the way that proteins interact.

*Joint work with David Sontag

2 Likelihood

We discuss two different likelihood models for the data.

2.1 Local Normalization

For each pair of proteins, we assume that there are three possible events which may trigger sufficient signs of interaction: (1) true PPI, (2) self-interaction of protein #1, and (3) self-interaction of protein #2. Thus, our model of an experiment yielding a negative result for a particular protein pair is simply a product of the individual probabilities of non-events:

$$P(Y_{ij} = 0|\theta_{ij}, \phi_i, \phi_j) = \theta_{ij}\phi_i\phi_j, \tag{1}$$

where $\theta_{ij} \in [0, 1]$ is the chance that proteins i and j do not interact, $\phi_i \in [0, 1]$ ($\phi_j \in [0, 1]$) is the chance that protein i (j) does not self-interact.

This model has one significant drawback: local normalization. In effect, we are assuming complete independence between the three (non-)events.

2.2 Global Normalization

We can rectify this by removing normalization from the individual variables; for mathematical convenience, we switch to exponential parameters. Our new unnormalized likelihood is nearly identical to the likelihood of the "local" model,

$$P(Y_{ij} = 0|\alpha_{ij}, \beta_i, \beta_j) \propto e^{\alpha_{ij}} e^{\beta_i} e^{\beta_j} \tag{2}$$

except that now elements of the product range along the non-negative reals: $e^{\alpha_{ij}}, e^{-\beta_i}, e^{-\beta_j} \in \mathbb{R}_+$. There are only two possible outcomes (for each Y_{ij}), so normalization of the model is trivial:

$$P(Y_{ij} = 0|\alpha_{ij}, \beta_i, \beta_j) = \frac{e^{\alpha_{ij}+\beta_i+\beta_j}}{1 + e^{\alpha_{ij}+\beta_i+\beta_j}}. \tag{3}$$

This is also known as a "logistic" or "logistic regression" model.

2.2.1 Logistic Regression

Logistic Regression is a binary classification model. Typically, there are a number of data points, represented as real-valued feature vectors, $x_1, \ldots, x_n \in \mathbb{R}^d$, and corresponding binary labels, $y_1, \ldots, y_n \in \{-1, +1\}$. The model learns a weight vector, $w \in \mathbb{R}^d$ and threshold, $\theta \in \mathbb{R}$, to maximize the likelihood (or posterior) of the data,

$$P(y_1, \ldots, y_n|w, \theta; x_1, \ldots, x_n) = \prod_i \frac{e^{x_i^T w - \theta}}{1 + e^{x_i^T w - \theta}}. \tag{4}$$

The threshold, θ, defines the decision boundary once points have been mapped to $\mathbb{R}$ via dot-product with the weight vector, w.

In our “global” or “logistic” model of protein interaction, the PPI parameter, α_{ij}, corresponds to the dot-product between feature and weight vectors; the negative sum of self-interaction variables, $\beta_i + \beta_j$, corresponds to the threshold.

Note that Logistic Regression corresponds to merely one of a number of different loss function for binary classification (Rennie & Srebro, 2005). We can easily substitute other loss functions to yield alternate classification models here.

3 Prior

Though we have one parmeter for each protein pair (constituting a matrix of parameters), it likely that relatively few factors determine the result of the experiment(s). A common approach is to limit the rank of the matrix of parameters. However, this leads to a non-convex optimization problem with many local minima. An alternative approach utilized by Srebro et al. (2005) is to use the trace norm of a matrix to regularize the complexity of the parameter matrix. Utilization of the trace norm in this way yields a convex optimizaiton problem with a low-rank solution.

We can incorporate the trace norm into a prior on the matrix of protein pair parameters. We achieve the effect attained by Srebro et al. (2005) by using a prior which is proportional to the exponentiated negative trace norm of the parameter matrix,

$$P(\alpha|\lambda) \propto \exp(-\lambda\|\alpha\|_\Sigma), \tag{5}$$

where λ controls the weighting of the prior, and $\|\alpha\|_\Sigma$ designates the trace norm of the parameter matrix α.

Since there is little worry of overfitting with respect to the self-interaction parameters, we assume an uninformative prior on β, $P(\beta) \propto 1$.

4 Model

Utilizing the “global” likelihood and the trace norm prior, we achieve a joint model,

$$P(Y, \alpha, \beta|\lambda) = \prod_{i,j} P(Y_{ij}|\alpha_{ij}, \beta_i, \beta_j)P(\alpha|\lambda)P(\beta). \tag{6}$$

Maximization of this probability with respect to α and β corresponds to maximum a posteriori, which is a standard way of infering parameter settings. Note that there is no need to normalize the trace norm.

References

Rennie, J. D. M., & Srebro, N. (2005). Loss functions for preference levels: Regression with discrete ordered labels. *Proceedings of the IJCAI Multidisciplinary Workshop on Advances in Preference Handling*.

Sontag, D., Singh, R., & Berger, B. (2007). Probabilistic modeling of systematic errors in two-hybrid experiments. *Proceedings of the Pacific Symposium on Biocomputing*.

Srebro, N., Rennie, J. D. M., & Jaakkola, T. S. (2005). Maximum margin matrix factorization. *Advances in Neural Information Processing Systems 17*.