Notes

Jackson Loper

January 2019

1 Objective

For any fixed matrix X with entries $X_{cg} \in \{0,1\}$ and any $n \geq 1$, let

$$L = \sum_{c,g} \left((X_{c,g} - .5) \left(\sum_k^n Z_{ck}\alpha_{gk} \right) - \log 2\cosh \frac{1}{2} \sum_k^n Z_{ck}\alpha_{gk} \right)$$

We here consider the problem of maximizing L with respect to α, z.

We additionally consider the case that we would like to maximize a regularized objective. Specifically, let

$$R^\alpha = \sum_g -\frac{1}{2}\alpha_g^T D_g^\alpha \alpha_g + \alpha_g^T d_g^\alpha$$
$$R^z = \sum_c -\frac{1}{2} z_c^T D_c^z z_c + z_c^T d_c^z$$

where for each g we have D_g^α is an $n \times n$ square matrix, d^α is a n-vector, and likewise for D^z, d^z. We can incorporate these regularizations by trying to maximize $L + R^\alpha + R^z$ instead.

2 What this code provides

1. $z, \alpha \leftarrow$ `logistic_svd.numpy_version.initialize`(X). Given X, uses SVD to give a reasonable initial estimate for z, α.

2. $\alpha' \leftarrow$ `logistic_svd.numpy_version.update_alpha`$(X, z, \alpha, D^\alpha, d^\alpha)$. Given X, D^α, d^α and an initial guess z, α, this function calculates an improved estimate for α', i.e. $L(z,\alpha) + R^\alpha(\alpha) \leq L(z,\alpha') + R^\alpha(\alpha')$. Note that, by the symmetry of this problem, this can be used to update z as well.

3. $\alpha' \leftarrow$ `logistic_svd.torch_version.update_alpha`$(X, z, \alpha, D^\alpha, d^\alpha)$. Same as above, but taking torch tensors as input instead of numpy arrays.

4. $L \leftarrow$ `logistic_svd.numpy_version.logistic_likelihood`(X, z, α). Calculates the (unregularized) objective.

5. $L \leftarrow$ `logistic_svd.torch_version.logistic_likelihood`(X, z, α). Same but for torch.

6. $L \leftarrow$ `logistic_svd.numpy_version.quadratic`(z, D^z, d^z). Calculates the regularization.

7. $L \leftarrow$ `logistic_svd.torch_version.quadratic`(z, D^z, d^z). Same but for torch.

3 How the updates work: minorization

Observe that for any initial condition, $\tilde{Z}, \tilde{\alpha}$, we may obtain a simple minorizaton for this problem. Indeed, let

$$
\begin{aligned}
M_{cg} = M_{cg}(\tilde{Z}, \tilde{\alpha}) &= \frac{\tanh\left(\frac{1}{2}\sum_k \tilde{Z}_{ck}\tilde{\alpha}_{gk}\right)}{2\sum_k \tilde{Z}_{ck}\tilde{\alpha}_{gk}} \\
\kappa_{cg} = \kappa_{cg}(\tilde{Z}, \tilde{\alpha}) &= \frac{1}{2}M_{cg}\left(\sum_k \tilde{Z}_{ck}\tilde{\alpha}_{gk}\right)^2 - \log 2\cosh\frac{1}{2}\sum_k \tilde{Z}_{ck}\tilde{\alpha}_{gk} \\
\tilde{L}_{M,k}(Z, \alpha) &= \sum_{c,g}\left(X_{c,g}\left(\sum_k Z_{ck}\alpha_{gk}\right) + \kappa_{cg} - \frac{1}{2}M_{cg}\left(\sum_k Z_{ck}\alpha_{gk}\right)^2\right)
\end{aligned}
$$

Then observe that

$$\tilde{L}_{M,k}(\tilde{Z}, \tilde{\alpha}) = L(\tilde{Z}, \tilde{\alpha})$$

Furthermore, it is well-known that

$$\tilde{L}_{M,k}(Z, \alpha) \leq L(Z, \alpha) \qquad \forall Z, \alpha$$

Thus $\tilde{L}$ is a so-called "minorizer" for L from the initial condition $\tilde{Z}, \tilde{\alpha}$. We can therefore be guaranteed that if we can find Z, α that improves $\tilde{L}$, it will also improve our value of L. That is, if we can find Z, α such that $\tilde{L}_{M,k}(Z, \alpha) > \tilde{L}_{M,k}(\tilde{Z}, \tilde{\alpha})$, then we will also have $L(Z, \alpha) > L(\tilde{Z}, \tilde{\alpha})$. This suggests the following iterative process:

1. Start with some initial condition $\tilde{Z}, \tilde{\alpha}$.
2. Calculate $M(\tilde{Z}, \tilde{\alpha}), k(\tilde{Z}, \tilde{\alpha})$
3. Find Z, α such that $\tilde{L}_{M,k}(Z, \alpha) > \tilde{L}_{M,k}(\tilde{Z}, \tilde{\alpha})$
4. Set $\tilde{Z} \leftarrow Z$, $\tilde{\alpha} \leftarrow \alpha$, go to step 2.

To enact this procedure, the key difficulty is step 3. That is, we need to be able to make progress on the surrogate problem $\tilde{L}$. It is to this problem we now turn our attention.

4 Progress on the surrogate problem $\tilde{L}$

Here we consider the problem of optimizing

$$\tilde{L}_{M,k}(Z,\alpha) = \sum_{c,g}\left(X_{c,g}\left(\sum_k Z_{ck}\alpha_{gk}\right) - \frac{1}{2}M_{cg}\left(\sum_k Z_{ck}\alpha_{gk}\right)^2\right)$$

Note we have dropped the κs that appeared in the previous section, since it is constant with respect to our objects of interest.

This problem can be optimized via coordinate ascent, alternating between Z and α. For example, let us consider only the case that we fix α and try to optimize Z. Note that with α fixed the problem is now separable over the cs. In particular, dropping constants, we see that for each c separately we need to optimize a problem of the form

$$f_c(z_c) = \sum_{g}\left(X_{c,g}\left(\sum_k Z_{ck}\alpha_{gk}\right) - \frac{1}{2}M_{cg}\left(\sum_k Z_{ck}\alpha_{gk}\right)^2\right)$$

Take derivatives:

$$\frac{\partial}{\partial z_{ck}} f_c(z_c) = \sum_g X_{c,g}\alpha_{gk} - M_{cg}\alpha_{gk}\left(\sum_{k'} Z_{ck'}\alpha_{gk'}\right)$$

Setting equal to zero, we see that the optimal α_g will be achieved by taking

$$\Gamma_{k,k'} = \sum_g M_{cg}\alpha_{gk}\alpha_{gk'}$$
$$z_c^* = \Gamma^{-1}\alpha^T X_c$$

We can do the same kind of update for α.

5 Initialization

If we initialize our problem with $Z = \alpha = 0$, our first minorization is given by taking $M_{cg} = \lim_{\epsilon\to 0}\tanh(\epsilon/2)/(2\epsilon) = .25$. This leads to the surrogate problem

$$\tilde{L}_{M,k}(Z,\alpha) = \sum_{c,g}\left(X_{c,g}\left(\sum_k Z_{ck}\alpha_{gk}\right) - \frac{1}{8}\left(\sum_k Z_{ck}\alpha_{gk}\right)^2\right)$$

It is easy to see that this problem is solved by taking Z, α as the first left and right singular vectors of $4(X - .5)$, each multiplied by the square root of the corresponding singular value. This gives a good initialization.

6 Regularization

Introducing per-c and per-g quadratic regularizations is straightforward. WLOG, let us consider updating z_c. After the minorization recall that the problem has become separable. The objective for a particular c, with regularization, is then

$$f_c(z_c) = -\frac{1}{2} z_c^T D z_c + d^T z_c + \sum_g \left(X_{c,g} \left(\sum_k Z_{ck} \alpha_{gk} \right) - \frac{1}{2} M_{cg} \left(\sum_k Z_{ck} \alpha_{gk} \right)^2 \right)$$

It is straightforward to see that this leads to the updates

$$\begin{aligned} \Gamma_{k,k'} &= \sum_g M_{cg} \alpha_{gk} \alpha_{gk'} \\ z_c^* &= (\Gamma + D)^{-1} (\alpha^T X_c + d) \end{aligned}$$