Statistical Optimization of Non-Negative Matrix Factorization

Anoop Korattikara* Levi Boyles* Max Welling* Jingu Kim† Haesun Park†

*University of California, Irvine †Georgia Institute of Technology

Summary

Non-Negative Matrix Factorization (NMF) is a dimensionality reduction technique that is useful for obtaining a “sum-of-parts” decomposition of data. We describe an optimization method that utilizes the stochastic nature of the data to highly speed up a state of the art NMF algorithm: the Block Principal Pivoting (BPP) method.

Non-Negative Matrix Factorization

Alternating Non-Negative Least Squares (ANLS):

\[ \min_{W \geq 0} \| W H - Y \|_F^2 \]

subject to \( \forall i, W_i, H_i \geq 0 \)

Update is asymptotically normal: If we assume

1) \( Q_u = E[cc^T] \) is positive definite
2) \( E[|c|^2] = 0 \)
3) \( \text{Var}[|c|^2] = \sigma^2 \)

Then \( x \sim \mathcal{N}(\mu, \Sigma) \) where \( \Sigma = \frac{Q_u}{\sigma^2} \). Note: variance inversely proportional to N

Statistical Optimization Idea

- Initial iterations / far from optima
- \( \mu \), \( \Sigma \)
- Principled stopping criterion,

\[ \min_{x} x = (C^T C)^{-1} C^T b \]

This is the ordinary least squares (OLS) Estimator:

Stochastic nature of data – Consider \((c,b)\) as i.i.d instances of RVs \((c,b)\).

We have the stochastic generative process:

\[ b = C^T x + \epsilon \]

Experimental Results

High speed-up over BPP on 3 real world datasets

BPP method for NNLS problems

Kim & Park (2008)

1. Partition index set \([1,…d]\) into sets \( F \& G \)
2. Set \( x_F, y_G = 0 \) (for satisfying complementary slackness)
3. Solve \( x_F = (C_F^T C_F)^{-1} C_F^T b \) and \( y_G = (C_G^T C_G)^{-1} x_F - C_G^T b \)
4. If \( x_F \geq 0 \) and \( y_G \geq 0 \), then \( x = (x_F, y_G) \) is an optimal solution. Done!

Else exchange indices between \( F & G \) and goto 2

Hypothesis Testing

- Test: Proposed update direction is within 90° of the true update direction.
- \( p_i \): is the probability that our update direction is wrong and should be small.

Conclusions

Advantages:
- High speed-up over BPP
- Single parameter to tune

Drawbacks:
- Principled stopping criterion, avoids over-fitting
- General method, not limited to NMF

Unpredictable if the normality assumptions do not hold e.g. because of sparsity in design matrix, multiplication of random variables etc.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No’s. 0447903, 0914783.