A spectral approximation to the Indian Buffet Process

James Robert Lloyd, Zoubin Ghahramani
Department of Engineering
Cambridge University

Abstract

Can we use spectral methods to get a fast model based on the IBP?

1 An IBP model

Consider a simple linear Gaussian model of the form

$$\begin{aligned} A &\sim \mathcal{N}(0, \sigma_A^2 I) & (1.1)\\ Z &\sim \text{IBP}(\alpha) & (1.2)\\ X &\sim \mathcal{N}(ZA, \sigma_X^2 I). & (1.3) \end{aligned}$$

Copying from [2] we should be guided by terms such as

$$\begin{aligned} p(X|Z,A) &\propto \exp(-\text{tr}((X-ZA)^{\mathrm{T}}(X-ZA))) & (1.4)\\ p(X|Z) &\propto |Z^{\mathrm{T}}Z + \frac{\sigma_X^2}{\sigma_A^2}I|^{D/2}\exp(-\text{tr}(X^{\mathrm{T}}(I - Z(Z^{\mathrm{T}}Z + \frac{\sigma_X^2}{\sigma_A^2}I)^{-1}Z^{\mathrm{T}})X)). & (1.5) \end{aligned}$$

For simplicity, we might initially consider simpler priors on Z and or a maximum likelihood framework.

2 Spectral clustering notes

For a similarity matrix G, with adjacency matrix W and diagonal degree matrix D, the unnormalized graph Laplacian is defined as

$$L = D - W \tag{2.1}$$

has the following property [3]

$$f'Lf = \frac{1}{2}\sum w_{ij}(f_i - f_j)^2 \tag{2.2}$$

and the multiplicity of the zero eigenvalue of L is equal to the number of connected components in G.

Spectral clustering is approximately solving the min cut problem for the weighted adjacency matrix i.e.

$$\operatorname*{argmin}_{Z_1} = W(Z_1, \bar{Z}_1) := \sum_{i \in Z_1, j \in \bar{Z}_1} w_{ij} \tag{2.3}$$

3 Relations between the two

Suppose we are trying to estimate A and Z by maximum likelihood. In particular, consider estimating the kth column of Z and the corresponding row of A keeping all other parameters fixed. Our objective can be stated as trying to minimise

$$||X - Z_{-k}A_{-k} - Z_kA_k|| \tag{3.1}$$

where $||.||$ is some distance metric on matrices (i.e. the appropriate one to make this equivalent to maximum likelihood).

Let $\tilde{X} = X - Z_{-k}A_{-k}$. For a given Z_k, the maximum likelihood estimation of A is equivalent to minimising

$$\sum_{i \in Z_k} |\tilde{x}_i - \beta_k| + \sum_{i \in \bar{Z}_k} |\tilde{x}_i - \bar{\beta}_k| \tag{3.2}$$

over β_k and $\bar{\beta}_k$. We can now create a link to spectral clustering.

Consider the following constant

$$\begin{aligned} C &= \sum_{i,j} |\tilde{x}_i - \tilde{x}_j| && (3.3) \\ &= \sum_{i,j \in Z_k} |\tilde{x}_i - \tilde{x}_j| + \sum_{i,j \in \bar{Z}_k} |\tilde{x}_i - \tilde{x}_j| + 2 \sum_{i \in Z_k, j \in \bar{Z}_k} |\tilde{x}_i - \tilde{x}_j| && (3.4) \end{aligned}$$

and then consider maximising $\sum_{i \in Z_1, j \in \bar{Z}_1} |\tilde{x}_i - \tilde{x}_j|$ over Z_k. This can be recast as minimising quantities of the form $\sum_{i \in Z_k, j \in \bar{Z}_k} (1 - \alpha|\tilde{x}_i - \tilde{x}_j|)$ for any $\alpha > 0$. For small enough α all summands will be positive and this can be phrased as a min cut problem i.e. approximate spectral clustering.

Thus, spectral clustering is approximately equivalent to minimising $\sum_{i,j \in Z_k} |\tilde{x}_i - \tilde{x}_j| + \sum_{i,j \in \bar{Z}_k} |\tilde{x}_i - \tilde{x}_j|$.

Let $\hat{\beta}_k = \operatorname{argmin}_\beta \sum_{i \in Z_k} |\tilde{x}_i - \beta|$. Using this definition and the triangle inequality, we get the following

$$|Z_k| \sum_{i \in Z_1} |\tilde{x}_i - \hat{\beta}| \leq \sum_{i,j \in Z_1} |\tilde{x}_i - \tilde{x}_j| \leq 2|Z_k| \sum_{i \in Z_1} |\tilde{x}_i - \hat{\beta}_k|. \tag{3.5}$$

i.e. we can show that min cut is optimising a bound on the maximum likelihood objective.

The bound is only tight when the problem is degenerate i.e. this is not yet a guarantee, just a heuristic. The tightness of these bounds could be demonstrated in a probabilistic sense by assuming the data was generated by a linear binary model.

4 A natural iterative algorithm

In the maximum likelihood setting this is easy, the following algorithm can be justified using the arguments above.

- Find a binary clustering, Z_1, using spectral clustering applied to X
- Fit maximum likelihood parameters to yield $\hat{A}_1$
- Obtain a new clustering, Z_2, by applying spectral clustering to $X - Z_1 A_1$. This is a sort of iterative conditional maximisation algorithm
- Fit maximum likelihood parameters to yield $\hat{A}_{1:2}$
- Obtain a new clustering, Z_3, by applying spectral clustering to $X - Z_{1:2}A_{1:2}$. This is a sort of iterative conditional maximisation algorithm
- et cetera... with looping, split merge equivalents and other nice things developed for samplers.

5 Next steps / questions

- Can we modify the above argument to be more Bayesian?
- In doing so, can we look into the min cut / spectral clustering approximation to find an appropriate graph Laplacian?
- How different is this to orthogonal projections [1]?
- Try it? Choosing sensible parameters for the spectral clustering will not be entirely easy.
- Can I make a guarantee about improving the marginal likelihood? Possibly not?

6 Some relevant literature

[2] [3] [1] [4] [5] [6]

References

[1] Cui, Y., Fern, X. Z., and Dy, J. G. (2007). Non-redundant Multi-view Clustering via Orthogonalization. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, **3**, 133–142.

[2] Griffiths, T. L. and Ghahramani, Z. (2011). The Indian Buffet Process : An Introduction and Review. *Journal of Machine Learning Research (JMLR)*, **12**, 1185–1224.

[3] Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, **17**(4), 395–416.

[4] Niu, D., Dy, J. G., and Jordan, M. I. (2010). Multiple Non-Redundant Spectral Clustering Views. In *Proceedings of the International Conference on Machine Learning (ICML_.*

[5] Niu, D., Dy, J. G., and Jordan, M. I. (2011). Dimensionality Reduction for Spectral Clustering. In *Proceeding of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15.

[6] Niu, D., Dy, J. G., and Ghahramani, Z. (2012). A Nonparametric Bayesian Model for Multiple Clustering with Overlapping Feature Views. In *Proceedings of the International Conference on Artifical Intelligence and Statistics (AISTATS)*.