laplacian_smoothing

December 13, 2017

1 Laplacian smoothing

By `@jhmarcus`

Here I explore the possible relationship beteween laplacian smoothing and the autoregressive models I've been looking into. Something is wrong with formatting here (need to figure it out) . . .

2 Background

Laplacian smoothing is a way to smooth out observed values on the nodes of the graph see

- http://www.stat.cmu.edu/~cshalizi/networks/16-1/lectures/13/lecture-13.pdf
- http://www.stat.cmu.edu/~ryantibs/papers/sparsify.pdf

It can be conceived as the following optimization problem . . .

$$\underset{\tilde{\mathbf{y}} \in \mathbb{R}^n}{min} \; ||\mathbf{y} - \tilde{\mathbf{y}}||_2^2 + \lambda \sum_{i,j} w_{ij}(\tilde{y}_i - \tilde{y}_j)^2$$

Where w_{ij} is the edge weight. We can represent this in matrix form as . . .

$$\underset{\tilde{\mathbf{y}} \in \mathbb{R}^n}{min} \; ||\mathbf{y} - \tilde{\mathbf{y}}||_2^2 + \lambda \tilde{\mathbf{y}}^T \mathbf{L} \tilde{\mathbf{y}}$$

where $\mathbf{L}$ is the graph laplacian. It can be shown the solution to this problem is analytical . . .

$$\tilde{\mathbf{y}}^* = (\mathbf{I} + \lambda \mathbf{L})^{-1} \mathbf{y}$$

$$\mathbf{y} = (\mathbf{I} + \lambda \mathbf{L}) \tilde{\mathbf{y}}^*$$

Lets now use this to motivate the setup an autoregressive process . . .

$$\mathbf{y} = (\mathbf{I} + \lambda \mathbf{L}) \mathbf{y} + \nu$$

where $\nu \sim \mathcal{N}(\nu | \mathbf{0}, \sigma^2 \mathbf{I})$ . . .

$$\mathbf{y} - (\mathbf{I} + \lambda \mathbf{L}) \mathbf{y} = \nu$$

$$\mathbf{y}\big(\mathbf{I} - (\mathbf{I} + \lambda \mathbf{L})\big) = \nu$$

$$\mathbf{y} = \frac{1}{\lambda}\mathbf{L}^{-1}\nu$$

This implies that ...

$$\mathbf{y}|\lambda, \sigma^2, \mathbf{L} \sim \mathcal{N}\Big(\mathbf{y}|\mathbf{0}, \frac{\sigma^2}{\lambda}(\mathbf{L}\mathbf{L}^T)^{-1}\Big)$$

Because λ and σ^2 are unidentifiable let $\tau = \frac{\sigma^2}{\lambda}$ ...

$$\mathbf{y}|\tau, \mathbf{L} \sim \mathcal{N}\big(\mathbf{y}|\mathbf{0}, \tau(\mathbf{L}\mathbf{L}^T)^{-1}\big)$$

This is exactly the covariance derived in Hanks 2016 from the perspective of a spatio-temporal random-walk. Perhaps this explains why it performs the best in the isolation_by_srw notebook? For more intution lets consider a two deme example (assuming $\lambda = 1$) ...

`todo`