Laplacian smoothing

By `@jhmarcus`

Here I explore the possible relationship between laplacian smoothing and the autoregressive models I've been looking into.

Background

Optimization perspective

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 with a penalized least-square loss ...

$$\mathcal{L}(\tilde{\mathbf{y}}) = \underset{\tilde{\mathbf{y}}}{min} \; ||\mathbf{y} - \tilde{\mathbf{y}}||_2^2 + \lambda \sum_{i,j} m_{ij}(\tilde{y}_i - \tilde{y}_j)^2$$

Where m_{ij} is the edge weight (migration rate which are assumed to be fixed in this case) between nodes i and j, $\mathbf{y}$ is the observed data (mean-centered genotypes at a single SNP observed in n people over geographic space), $\tilde{\mathbf{y}}$ is the unobserved smooth underlying function on the nodes of the graph, and λ is a smoothness parameter which penalizes un-smooth functions. We can represent this in matrix form as ...

$$\mathcal{L}(\tilde{\mathbf{y}}) = \underset{\tilde{\mathbf{y}}}{min} \; ||\mathbf{y} - \tilde{\mathbf{y}}||_2^2 + \lambda \tilde{\mathbf{y}}^T \mathbf{L} \tilde{\mathbf{y}}$$

where $\mathbf{L}$ is the graph laplacian. If we differentiate with respect to $\tilde{\mathbf{y}}$ and solve for when the loss is 0 it can be shown that the $\tilde{\mathbf{y}}$ that minimizes the loss is ...

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

Solving for $\mathbf{y}$ we can see that ...

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

Autoregressive process perspective

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

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

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

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

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

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

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 spatial-random-walks notebook?

Some intuition

For more intuition lets consider a two deme example (assuming $\lambda = 1$) ...

$$\Big(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} w_{12} & -w_{12} \\ -w_{21} & w_{21} \end{bmatrix} \Big) \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} =$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} w_{12}y_1 - w_{12}y_2 \\ w_{21}y_2 - w_{21}y_1 \end{bmatrix}$$

Lets let the edge weights be symmetric so that $\alpha = w_{12} = w_{21}$ ...

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \alpha y_1 - \alpha y_2 \\ \alpha y_2 - \alpha y_1 \end{bmatrix} = \begin{bmatrix} y_1 + \alpha(y_1 - y_2) \\ y_2 + \alpha(y_2 - y_1) \end{bmatrix}$$

Thus we can see if α large and the distance between values on the two nodes are large they will be smoothed towards each other whereas if α is small they will be less smooth. Of course the bigger λ is the more smooth the data will be ... its effectively making the edge weights larger.

Circuit theory

TODO: *expand upon this interpretation*

We can interpret ν_i as random variable representing the total amount of current entering the circuit through i and such $\mathbf{y}_i$ is the potential at node i.