April 7, 2020

Before: "unsupervised" learning on graphs

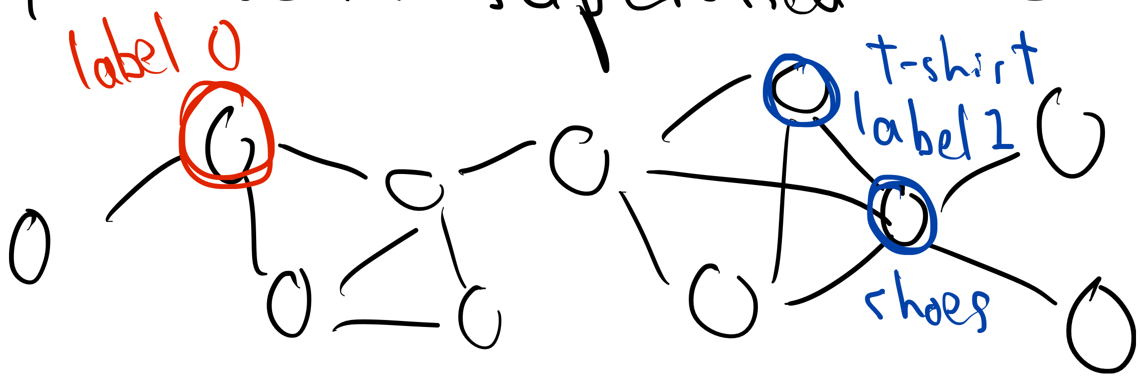
$$L = D - A \qquad N = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$$

$$\left.\begin{aligned} Lx &= \lambda_2 x \\ Nx &= \lambda_2 x \end{aligned}\right\} \quad \text{clusters based on sign, "sweep cut"}$$

$$\left.\begin{aligned} L X_k &= \Lambda_k X_k \\ N X_k &= \Lambda_k X_k \end{aligned}\right\} \quad n \; \overset{k}{\boxed{\;X_k\;}} \quad \text{node} \to \text{point in } \mathbb{R}^k$$

cuts, volumes as motivation

Today: "semi-supervised" learning



$G = (V, E)$

$H \subseteq V$ labelled 0/1

Goal: For $U = V \setminus H$ assign 0/1

Remember: $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$

$$\min_{x \in \mathbb{R}^n} \; \underbrace{x^T L x}_{\text{smoothness}}$$

s.t. $x_h = \text{label}(x_h), \; h \in H$

$\text{label}(x_h) \to \ell_h$

(Zhu, Gharamani, Lafferty '03)

Lagrangian: $\mathcal{L}(x, \lambda) = x^T L x - \sum_{h \in H} \lambda_h (x_h - l_h)$

$$\nabla \mathcal{L}(x^*, \lambda^*) = 0$$

$$\nabla_{x_u} \mathcal{L}(x, \lambda) = 2(Lx)_u$$

$$= 2[(Dx)_u - (Ax)_u]$$

$$= 0 \Rightarrow d_u x_u - \sum_v A_{uv} x_v = 0$$

$$x_u = \frac{1}{d_u} \sum_{(u,v) \in E} x_v$$

Algorithm: repeatedly average neighbors (label propagation)

$$x_u^{(k+1)} = \frac{1}{d_u} \sum_{(u,v)\in E} x_v^{(k)} \qquad u \in U \qquad x_u^{(0)} = 0$$

$$x_h^{(k+1)} = l_h \qquad h \in H$$

Will this converge?

$$x_u^{(k+1)} = \left[(D^{-1}A)\, x^{(k)} \right]_u$$

$P = D^{-1}A$ (row stochastic)

$$= \begin{bmatrix} P_{UU} & P_{UH} \end{bmatrix} \begin{pmatrix} x_U^{(k)} \\ x_H^{(k)} \end{pmatrix}$$

$$P = \begin{bmatrix} P_{uu} & P_{uH} \\ P_{Hu} & P_{HH} \end{bmatrix}$$

Claim: $\|P\|_2 = 1$

$(\|P^T\|_2 = 1)$

$$P^T x = 1 x \implies \|P\|_2 \geq 1$$

rows of P are stochastic

$$Py = \lambda y \quad \lambda > 1 \qquad \underbrace{[Py]_i}_{\text{convex comb. of entries of } y} = \lambda y_i > y_i$$

can't be true y_{max}

$$\implies \|P\|_2 = \|P^T\|_2 = 1$$

$$\cancel{\|x_u^{(k+1)}\| = \left\| \left[\begin{bmatrix} P_{uu} & P_{uH} \\ P_{Hu} & P_{HH} \end{bmatrix} \begin{pmatrix} x_u^{(k)} \\ x_H^{(k)} \end{pmatrix} \right]_u \right\|}$$

nonnegative

nonnegative

$$\cancel{\leq \|P x^{(k)}\|}$$

$$\cancel{\leq \|P\| \, \|x^{(k)}\| = \|x^{(k)}\|}$$

$$x_u^{(0)} = 0$$

$$x^{(1)} = P_{uu} x^{(0)} + P_{ul} x_l = z$$

z

$$x^{(2)} = P_{uu} x^{(1)} + z = P_{uu} z + z$$

Claim: $x^{(k)} = \sum_{j=0}^{k-1} P_{uu}^j z$

Induction: $x^{(k+1)} = P_{uu} x^{(k)} + z$

$$= P_{uu} \sum_{j=0}^{k-1} P_{uu}^j z + z = \sum_{j=0}^{k} P_{uu}^j z$$

Converges to $\lim_{k \to \infty} \sum_{j=0}^{k} P_{uu}^j z = (I - P_{uu})^{-1} z$

(Need $\|P_{uu}\|_2 < 1$ for limit to exist)

Have: $\|P_{uu}\|_2 \leq 1$

Suppose $P_{uu}^T z = z$, $z > 0$ by Perron-Frobenius (connected graph)

Choose $\|z\|_1 = 1$

$1^T P_{uu}^T z = z = 1 \Rightarrow 1^T P_{uu}^T = 1^T$

but $P_{uH} \neq 0$ and P is stochastic

$\Rightarrow$ some column of P_{uu}^T is strictly substochastic

$\Rightarrow$ $1^T P_{uu}^T \neq 1^T$ (contradiction)

Converges to what?

- $x_u^* = P_{uu} x_u^* + P_{uH} x_H$

$(I - P_{uu}) x_u^* = P_{uH} x_H \qquad x_u^* = (I - P_{uu})^{-1} P_{uH} x_H$

Can prove:

x_u^* = Prob(random walk starting at node u hits a "1" before it hits a "0")

Clustering interpretation

$$\min_{x} \; x^T L x = \sum_{(i,j)\in E} (x_i - x_j)^2 = \mathrm{cut}(V)$$

$$V = \{ i \mid x_i = 1 \}$$

s.t. $x_H = l_H$ $\quad \cancel{x_i \in \{0,1\}}$

soft clustering

solvable?

Multiple clusters?

$$\min_{X \in \mathbb{R}^{n\times k}} \mathrm{tr}(X^T L X)$$

soft

s.t. $\cancel{X_{il} \in \begin{cases} 1 & \text{if } i \text{ is in cluster } l \in \{1,\dots,k\} \\ 0 & \text{o/w} \end{cases}}$

$$\begin{bmatrix} 0010 \\ 1000 \end{bmatrix} \qquad X_{Hl} \; \{ \; \text{fixed}$$

$$\operatorname{tr}(X^T L X) = \sum_{\ell=1}^{k} x_\ell^T L x_\ell$$

$$\begin{array}{r} \text{label 3} \\ \text{label 2} \\ \\ \end{array} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ & & \vdots & \end{bmatrix}$$

k soft labelings (one for each class)

$$\text{label } i = \arg\max_{\ell} X_{i\ell}$$

---

min "unsmoothness" $f(x)$

s.t. labels we know: $x_H = \ell_H$

$$\min f(x) + \lambda \|x_H - \ell_H\|_2^2 \qquad 0/1$$

$x^T L x$ $\quad$ $x^T N x$ $\quad + \underbrace{\lambda \left\| \begin{pmatrix} x_U \\ x_H \end{pmatrix} - \begin{pmatrix} 0 \\ \ell_H \end{pmatrix} \right\|_2^2}_{\|x-y\|_2^2}$

$$\min \frac{1}{2}\left(x^T N x + \lambda \|x-y\|_2^2 \right)$$

$$\nabla_x = Nx + \lambda(x-y)$$

$$x^{(k+1)} = x^{(k)} - \alpha\left(Nx^{(k)} + \lambda\left(x^{(k)} - y\right)\right)$$

$\lambda = \frac{1-\alpha}{\alpha}$

$$= \alpha x^{(k)} - \alpha N x^{(k)} + (1-\alpha)y$$

$$= \alpha (I - N) x^{(k)} + (1-\alpha) y$$

$$= \alpha \underbrace{\left(D^{-1/2} A D^{-1/2}\right)}_{\text{global consistency}} x^{(k)} + \underbrace{(1-\alpha)y}_{\text{local consistency}}$$

$$\left(I - \alpha D^{-1/2} A D^{-1/2}\right) x^* = (1-\alpha)y$$

$\alpha \in (0,1)$

$$\underbrace{D^{-1/2} x^{(k+1)}}_{z^{(k+1)}} = \alpha \underset{D^{-1}}{\cancel{D^{-1/2}}} A \underbrace{D^{-1/2} x^{(k)}}_{z^{(k)}} + (1-\alpha) D^{-1/2} y$$

$$D^{-1} A = P$$

$$(Ax)_i = \sum_{(i,j)\in E} x_j \qquad (Px)_i = \frac{1}{d_i} \sum_{(i,j)} x_j$$

$$\left(D^{-1/2} A D^{-1/2} x\right)_i = \frac{1}{\sqrt{d_i}} \sum_{(i,j)} x_j / \sqrt{d_j}$$