Today:

Linear Algebra & random walks

Saving random bits via random walks

"Well, that's the news from Lake Wobegon, where all the women are strong, the men are good looking, and all the children are above average"

— Garrison Keillor "A Prairie Home Companion"

Linear Algebra Review

def. v is an eigenvector of A with corresponding eigenvalue λ iff $vA = \lambda v$

def $\mathcal{L}_2$-norm of $v = (v_1 \dots v_n) = \sqrt{\sum_{i=1}^{n} v_i^2}$

def. $v^{(1)} \dots v^{(m)}$ orthonormal if

$$\underbrace{v^{(i)} \cdot v^{(j)}}_{\text{inner product}} = \begin{cases} 1 & \text{if } i=j \quad \text{normal} \\ 0 & \text{o.w.} \quad \text{orthogonal} \end{cases}$$

inner product $\sum_{l} v^{(i)}(l) \cdot v^{(j)}(l)$

example! P = transition matrix of d-reg undirected graph (doubly stochastic)

$$\left(\tfrac{1}{n} \dots \tfrac{1}{n}\right) \cdot P = 1 \cdot \left(\tfrac{1}{n} \dots \tfrac{1}{n}\right)$$

also : $\left(\tfrac{1}{\sqrt{n}} \dots \tfrac{1}{\sqrt{n}}\right) \cdot P = 1 \cdot \underbrace{\left(\tfrac{1}{\sqrt{n}} \dots \tfrac{1}{\sqrt{n}}\right)}_{\mathcal{L}_2\text{-norm} = 1}$

$\mathcal{L}_2$-norm $= 1$
$\shortparallel$
normal

Just like Lake Wobegon, where all the children are above-average

"Important" Thm: transition matrix P real + symmetric
$\Rightarrow \exists$ e-vectors $v^{(1)} \dots v^{(n)}$ forming orthonormal basis with corresponding e-values $1 = \lambda_1 \geq |\lambda_2| \geq \dots \geq |\lambda_n|$

$+ \quad v^{(1)} = \frac{1}{\sqrt{n}} (1 \dots 1)$ ← set so that $\|v^{(1)}\|_2 = 1$

Assume P has all positive entries & has e-vecs $v^{(1)} \dots v^{(n)}$ with corresponding evals $\lambda_1 \dots \lambda_n$

Fact (1) αP has evecs $v^{(1)} \dots v^{(n)}$ with corresp evals $\alpha\lambda_1, \dots, \alpha\lambda_n$

(2) $P+I$ " " " " " " $\lambda_1+1, \dots, \lambda_n+1$

(3) P^k " " " " " " $\lambda_1^k, \dots, \lambda_n^k$

(4) P stochastic $\Rightarrow |\lambda_i| \leq 1 \quad \forall i$

why? (1) $vP = \lambda v \iff v\alpha P = \lambda \alpha P$

(2) $v(P+I) = vP + vI = \lambda v + v = (\lambda+1)v$ self-loops: $\frac{P+I}{2}$ = "stay put with prob ½ + walk with prob ½"

(3) $vP^k = (vP)P^{k-1} = \lambda v P^{k-1} = \lambda^2 v P^{k-2} = \dots = \lambda^k v$ K-step walks

(4) For all i, let $I = \{j \mid v_j^{(i)} > 0\}$

then $\lambda \sum_{j \in I} v_j^{(i)} = \sum_{j \in I} \sum_k v_k^{(i)} P_{kj}$

$$\leq \sum_{\substack{j,k \\ \text{st } j,k \in I}} v_k^{(i)} P_{kj}$$

$$\leq \sum_{k \in I} v_k^{(i)} \underbrace{\sum_{j \in I} P_{kj}}_{\substack{\leq 1 \\ \text{since} \\ \text{stochastic}}} \leq \sum_{k \in I} v_k^{(i)}$$

$\therefore \lambda \leq 1$

Note if $v^{(1)} \dots v^{(n)}$ orthonormal basis then

any vector w is expressible as linear combination of $v^{(i)}$'s

$$w = \sum \alpha_i v^{(i)}$$

+ L_2 norm of w is $\underbrace{\sqrt{\sum \alpha_i^2}}_{\text{why?}}$

$$\|w\|_2 = \sqrt{\sum \alpha_i v^{(i)} \cdot \sum \alpha_j v^{(j)}}$$

$$= \sqrt{\sum \alpha_i \alpha_j \underbrace{v^{(i)} \cdot v^{(j)}}_{\begin{cases} =0 & \text{if } i \neq j \\ =1 & \text{if } i = j \end{cases}}}$$

$$= \sqrt{\sum \alpha_i^2}$$

Mixing times

How long does it take to reach stationary distribution?

def. $\varepsilon > 0$

Mixing time, $T(\varepsilon)$, of M.C. A with stationary distribution π is min t s.t. $\forall\ \pi^{(0)}$, $\|\pi - \pi^{(0)} A^t\|_1 < \varepsilon$

def M.C. A is rapidly mixing if $T(\varepsilon) = \text{poly}(\log \underset{\uparrow \\ \#\text{ states}}{n}, \log 1/\varepsilon)$

examples: r.w. on complete graph, random graph

Thm P is transition matrix of undirected, nonbipartite,* d-regular connected graph

if bipartite, then $\lambda_2 = 1$

π_0 is start dist

π is stationary dist $= (\frac{1}{n}, \dots, \frac{1}{n})$ so $\pi P = \pi$

Then $\| \pi_0 P^t - \pi \|_2 \le |\lambda_2|^t$ ⇐ exponentially decreasing distance if $1-\lambda_2$ = constant

Proof

P real, symmetric $\Rightarrow$ e-vecs $v^{(1)} \dots v^{(n)}$ are orthonormal basis with evals $1 = \lambda_1 \ge |\lambda_2| \ge \dots \ge |\lambda_n|$

so any vector, in particular, π_0, can be expressed as linear combination of $v^{(i)}$'s

$$\pi_0 = \sum_{i=1}^{n} \alpha_i v^{(i)}$$

$$\text{so } \pi_0 \cdot P^t = \sum_{i=1}^{n} \alpha_i \underbrace{v^{(i)} P^t}_{= \lambda_i^t v^{(i)}}$$

$$= \underbrace{\alpha_1 \lambda_1^t}_{\lambda_1 \equiv 1} v^{(1)} + \alpha_2 \lambda_2^t v^{(2)} + \dots$$

then $\| \pi_0 P^t - \alpha_1 v^{(1)} \|_2 = \| \sum_{i=2}^{n} \alpha_i \lambda_i^t v^{(i)} \|_2$

$= \sqrt{\sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2t}}$ previous note

$\le |\lambda_2|^t \sqrt{\sum_{i=2}^{n} \alpha_i^2}$ since $|\lambda_2| \ge |\lambda_3| \ge \dots$

$\le |\lambda_2|^t \, \|\pi_0\|_2$ by Note on previous page + since $\sum_{i=0}^{1} \alpha_i^2 > 0$

$\le |\lambda_2|^t$ since $L_2 \le L_1 = 1$

$|\lambda_2|^t$ goes to 0 so has to be stationary

Reducing Randomness

For decision problem L,

Let $\mathcal{A}$ be algorithm s.t. 1) $\forall x \in L$ $\Pr[\mathcal{A}(x)=1] \geq 99/100$ almost always correct

2) $\forall x \notin L$ $\Pr[\mathcal{A}(x)=0] = 1$ always correct

To get error $< 2^{-k}$:

Method:	# random bits used
1) run k times & output majority	$O(kr)$
2) use p.i. random bits	$O(k+r)$
3) today: use random walk on graph to choose random bits	$r + O(k)$

Plan:

- associate all (random) strings in $\{0,1\}^n$ with nodes of a graph G
- problem of picking a random string is now equivalent to problem of picking a random node ← easier?

  picking several random strings ⟹ picking several nodes ↖ easier?

  picking several strings, one of which is "good" ⟹ picking several nodes, one of which is "good" ↖ "easier"!

The graph G:

- constant degree d-regular, connected, nonbipartite
- transition matrix P for r.w. on G has $|\lambda_2| \leq \frac{1}{10}$
  π uniform since d-reg
- $\#_{\text{nodes}} = 2^r$ $\sim$ r random bits

The Algorithm:

- pick random start node $w \in \{0,1\}^r$ — r bits
- Repeat K times:

  $w \leftarrow$ random neighbor of w — $O(1)$ bits $\times K$

  run $\mathcal{A}(x)$ with w as random bits

  if $\mathcal{A}$ outputs "$x \in L$" then output "$x \in L$" + halt

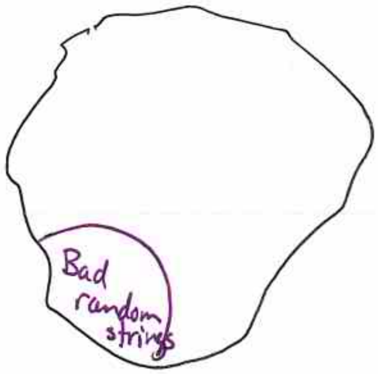
  else continue
- Output "$x \notin L$"

total random bits: $r + O(K)$

Claim: error of new algorithm $\leq \left(\frac{1}{5}\right)^K$ for $x \in L$
(still 0-error for $x \notin L$)

Behavior:

idea:



bad case - walk only on "bad" random strings & never get out to "good" random strings

why would this <u>not</u> work on arbitrary G?

e.g. G = line

if $x \notin L$: algorithm <u>never</u> errs (there are no bad strings)

if $x \in L$:

most random bits say $x \in L$: $\geq \frac{99}{100} \cdot 2^r$

define $B \leftarrow \{ w \mid \mathcal{A}(x)$ with random bits w is incorrect i.e. says $x \notin L \}$

"Bad w's"

$$|B| \leq \frac{2^r}{100}$$

Want linear algebraic way of describing walks that stay in badset:

define N diagonal matrix such that

$$N_w = \begin{cases} 1 & \text{if } w \in B \leftarrow \text{incorrect} \\ 0 & \text{o.w.} \leftarrow \text{correct} \end{cases}$$

Bad w's

$$N = \begin{pmatrix} 1,1, & & & & 0 \\ & 0_{0_{0_0}} & & & \\ & & 1_1 & & \\ & & & 0_{0_{0_0}} & \\ 0 & & & & 1_1 \end{pmatrix}$$

q any probability distribution

$\|qN\|_1 = \Pr_{w \in q}[w \text{ is bad}]$

i.e. pN deletes weight that finds a witness to $x \in L$

Can compose:

$\|q \cdot PN\|_1 = \Pr_{w \in q}$ [start at q, take a step & land on "bad"]

$\vdots$

$\|q \cdot (PN)^i\|_1 = \Pr_{w \in q}$ [start at q, take i steps + each is "bad"]

ignores whether start node is bad, this just hurts us so it is ok to ignore

Lemma $\forall \pi \quad \|\pi PN\|_2 \le \frac{1}{5}\|\pi\|_2$

First: How do we use the lemma?

If always see bad w's, then answer incorrect

$\Rightarrow \Pr[\text{incorrect}] \le \|p_0 \cdot (PN)^k\|_1$

$\le \sqrt{2^r}\,\|p_0 (PN)^k\|_2$ since $\|p\|_1 \le \sqrt{\text{domain size}} \cdot \|p\|_2$

$\le \cancel{\sqrt{2^r}} \cdot \underbrace{\|p_0\|_2}_{\cancel{\frac{1}{\sqrt{2^r}}}} \left(\frac{1}{5}\right)^k$ apply lemma k times

since start at uniform + L_2 norm of uniform $= \sqrt{\Sigma\left(\frac{1}{2^r}\right)^2} = \sqrt{\frac{1}{2^r}}$

$= \left(\frac{1}{5}\right)^k$

Proof of lemma let $V_1 \dots V_{2^n}$ be e-vects of P, + V_1 is s.t. $\|V_1\|_2 = 1$

note, $V_1 = \left(\frac{1}{\sqrt{2^n}}, \dots, \frac{1}{\sqrt{2^n}}\right)$

then $\pi = \sum_{i=1}^{2^n} \alpha_i V_i$

Note: 1) $\|\pi\|_2 = \sqrt{\alpha_i^2}$ (from before)

2) $\forall\ w \quad \|wN\|_2 = \sqrt{\sum_{i \in B} w_i^2} \leq \sqrt{\sum_i w_i^2} = \|w\|_2$

So:

$$\|\pi P N\|_2 = \left\| \sum_{i=1}^{2^n} \alpha_i v_i P N \right\|_2$$

$$= \left\| \sum_{i=1}^{2^n} \alpha_i \lambda_i v_i N \right\|_2$$

$$\leq \underbrace{\| \alpha_1 \lambda_1 v_1 N \|_2}_{A} + \underbrace{\left\| \sum_{i=2}^{2^n} \alpha_i \lambda_i V_i N \right\|_2}_{B}$$ Cauchy-Schwarz

bounding (A):

$$\|\alpha_1 \lambda_1 v_1 N\|_2 = \|\alpha_1 v_1 N\|_2$$ since $\lambda_1 = 1$

$$= |\alpha_1| \sqrt{\sum_{i \in B} \left(\frac{1}{\sqrt{2^n}}\right)^2}$$ since $V_1 = \left(\frac{1}{\sqrt{2^n}}, \dots, \frac{1}{\sqrt{2^n}}\right)$

use that uniform is unlikely to be on bad string

$$= |\alpha_1| \sqrt{\frac{|B|}{2^n}}$$

$$\leq \frac{|\alpha_1|}{10}$$ since $\frac{|B|}{2^n} \leq \frac{1}{100}$

$$\leq \frac{\|\pi\|_2}{10}$$ since $\|\pi\|_2 = \sqrt{\sum \alpha_i^2}$

Bounding (B) : $\| \sum_{i=2}^{2^r} \alpha_i \lambda_i v_i N \|_2 \le \| \sum_{i=2}^{2^r} \alpha_i \lambda_i v_i \|_2$ from note

$$= \sqrt{\sum (\alpha_i \cdot \lambda_i)^2}$$

use "mixing"

$$\le \sqrt{\sum \alpha_i^2 \left(\tfrac{1}{10}\right)^2} \qquad \lambda_i \le 1/10$$

$$\le \frac{1}{10} \|\pi\|_2$$

so: $\|\pi P N\|_2 \le \dfrac{\|\pi\|_2}{5}$ ∎