Lecture 10 Power Method

Sunday, September 25, 2016 9:03 AM

- Jenrich's algorithm is not very stable

  needs eigenvalue gaps in $(M_a M_b)^{-1}$ to be large.

- In practice, power method is more popular.

- Tensor $T = \sum_{i=1}^{r} \lambda_i a_i \otimes a_i \otimes a_i$

  a_i's are orthonormal. wlog $\lambda_i \geq 0$

  $u^{(0)} \sim N(0, \frac{1}{d} I)$

  for $i = 1$ to t

  $u^{(i)} = T(u^{(i-1)}, u^{(i-1)}, I)$

  $= \sum_{j=1}^{r} \lambda_j \langle u^{(i-1)}, a_j \rangle^2 a_j$

  (optional: $u^{(i)} = \frac{u^{(i)}}{\|u^{(i)}\|}$)

Theorem: With high probability, power method converges to a_i with largest $|\lambda_i \langle u^{(0)}, a_i \rangle|$ in $\log\log\frac{d}{\varepsilon} + \log d$ steps.

Proof: Let $u^{(i)} = \sum_{j=1}^{r} c_j^{(i)} a_j$

$$c_j^{(i)} = \lambda_j \left(c_j^{(i-1)}\right)^2$$

$$c_j^{(1)} = \lambda_j \left(c_j^{(0)}\right)^2$$

$$c_j^{(2)} = \lambda_j \left(\lambda_j \left(c_j^{(0)}\right)^2\right)^2 = \lambda_j^3 \left(c_j^{(0)}\right)^4$$

$$c_j^{(3)} = \lambda_j \left(\lambda_j^3 \left(c_j^{(0)}\right)^4\right)^2 = \lambda_j^7 \left(c_j^{(0)}\right)^8$$

$$C_j^{(i)} = \lambda_j^{2-1} (C_j^{(0)})^2$$

$$= (\lambda_j C_j^{(0)})^{2^i - 1} C_j^{(0)}$$

with high probability

$$\max |\lambda_j C_j^{(0)}| > \left(1 + \frac{1}{d^2}\right) \lambda_{j'} C_{j'}^{(0)} \quad (j' \neq \arg\max)$$

$$C_j^{(0)} \geq \frac{1}{d^2}$$

choose t such that $2^t - 1 \geq 3d^2 \log \frac{d}{\varepsilon}$

$$\frac{C_j^{(t)}}{C_{j'}^{(t)}} > \frac{\varepsilon}{d} \Longrightarrow \frac{u^{(t)}}{\|u^{(t)}\|} \text{ is } \varepsilon\text{-close to } a_j \quad \square$$

- Problem: $\{a_i\}$'s need to be orthogonal.
- Solution: Whitening.

$$T = \sum_{i=1}^{r} \lambda_i u_i \otimes u_i \otimes u_i$$

$$M = \sum_{i=1}^{r} \lambda_i u_i u_i^T$$

does not have to be the same, but need $\lambda_i > 0$

$$M = UDU^T$$

$$W = UD^{-\frac{1}{2}}$$

property: $W^T M W = I$

$$a_i = \lambda_i^{\frac{1}{2}} W^T u_i$$

then: $\|a_i\| = 1$, $\langle a_i, a_j \rangle = 0$

$$T(W, W, W) = \sum_{i=1}^{r} \lambda_i (W^T u_i) \otimes (W^T u_i) \otimes (W^T u_i)$$

$$= \sum_{i=1}^{r} \lambda_i^{-\frac{1}{2}} \left(\lambda_i^{\frac{1}{2}} W^T u_i\right)^{\otimes 3}$$

$$= \sum_{i=1}^{r} \lambda_i^{-\frac{1}{2}} a_i^{\otimes 3}$$

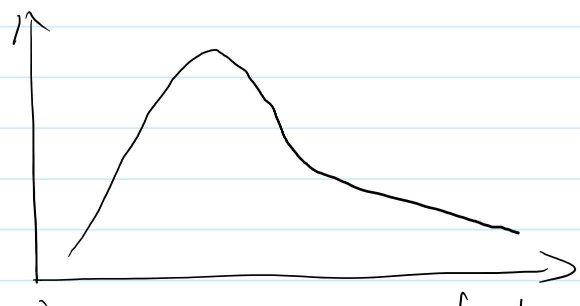
- How to find tensor structure?
  - Method of moments.
    - recall: model $D(\theta)$, want to estimate θ.
    - MoM: compute the moments of $x \sim D(\theta)$

      $$\mathbb{E}[x] \quad \mathbb{E}[xx^T] \quad \mathbb{E}[x \otimes x \otimes x] \quad \cdots$$

      then solve equations.
    - Example: $x \sim N(\mu, \sigma^2)$

      $$\mathbb{E}[x] = \mu \quad \mathbb{E}[x^2] = \mu^2 + \sigma^2 \quad \checkmark$$

    - identifiability: the equations have unique solution.
  - Pearson's Crabs
    - measures a certain ratio for crabs, distribution



Conjecture: mixture of two Gaussians.

Can compute $\mathbb{E}[x], \mathbb{E}[x^2] \cdots \mathbb{E}[x^6]$ and solve for parameters.

- Simple example: pure topic model

- Each document has only 1 topic

- $Q_{i,j} = \Pr[\text{1st word} = i, \text{2nd word} = j]$

$$= \sum_{l=1}^{r} \Pr[\text{topic} = l] \cdot A_{i,l} \cdot A_{j,l}$$

$$Q = \sum_{l=1}^{r} \Pr[\text{topic} = l] \, a_l a_l^T$$

- $T_{i,j,k} = \Pr[\text{1st word} = i, \text{2nd word} = j, \text{3rd word} = k]$

$$= \sum_{l=1}^{r} \Pr[\text{topic} = l] \, A_{i,l} A_{j,l} A_{k,l}$$

$$T = \sum_{l=1}^{r} \Pr[\text{topic} = l] \, a_l \otimes a_l \otimes a_l$$