Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Spectral Methods

Advanced Machine Learning for NLP

Jordan Boyd-Graber

TENSOR APPROACH

Big Idea

- You have a model
- What correlations should you see if model true
- Can you reverse the model from these correlations?

Big Idea

- You have a model
- What correlations should you see if model true
- Can you reverse the model from these correlations?
- Yes!

Simple Example: Mixture of Multinomials

Mixture of Multinomials

- k topics: $\phi_1, \dots \phi_k$
- Observe topic i with probability θ_i
- Observe m (exchangeable) words $w_1, \dots w_m$ iid from μ_i

- Given: m-word documents
- Goal: ϕ's, θ

Vector notation

- One-hot word encoding $w_1 = [0, 1, 0, \dots]^\top$
- ϕ_i are probability vectors
- Conditional probabilities are parameters

$$\text{Pr}[w_1] = \mathbb{E}[w_1 \,|\, \text{topic}\, i] = \phi_i \quad (1)$$

Method of Moments

- Find parameters consisten with observed moments
- Alternative to EM / objective-based techniques
- Topic model moments

$$\Pr[w_1] \tag{2}$$
$$\Pr[w_1, w_2] \tag{3}$$
$$\Pr[w_1, w_2, w_3] \tag{4}$$
$$\vdots \tag{5}$$

First Moment

With one word per document,

$$\Pr[w_1] = \sum_{i=1}^{k} \theta_i \phi_i \tag{6}$$

Not identifiable: only d numbers

Problem setup

- (Tensor) Want to find good solution to

$$T = \sum_{t=1}^{d} \theta_t \vec{\phi}_t \otimes \vec{\phi}_t \otimes \vec{\phi}_t \tag{7}$$

  - $\otimes$ is dyadic product, creating a $d \times d \times d$ matrix (similar to $d \times d$ Anchor term)

Problem setup

- (Tensor) Want to find good solution to

$$T = \sum_{t=1}^{d} \theta_t \vec{\phi}_t \otimes \vec{\phi}_t \otimes \vec{\phi}_t \tag{7}$$

  - $\otimes$ is dyadic product, creating a $d \times d \times d$ matrix (similar to $d \times d$ Anchor term)
- (Matrix) Want to find eigendecomposition of

$$M = \sum_{t=1}^{n} \theta_t \vec{\phi}_t \vec{\phi}_t^\top \tag{8}$$

- But we won't see actual M, it will have error $\mathscr{E}$

Problem setup

- (Tensor) Want to find good solution to

$$T = \sum_{t=1}^{d} \theta_t \vec{\phi}_t \otimes \vec{\phi}_t \otimes \vec{\phi}_t \tag{7}$$

  - $\otimes$ is dyadic product, creating a $d \times d \times d$ matrix (similar to $d \times d$ Anchor term)
- (Matrix) Want to find eigendecomposition of

$$M = \sum_{t=1}^{n} \theta_t \vec{\phi}_t \vec{\phi}_t^\top \tag{8}$$

- But we won't see actual M, it will have error $\mathcal{E}$
  - Unique if θ_i are
  - Solveable if $||\mathcal{E}||_2 < \min_{i \neq j} |\theta_i - \theta_j|$

Power iteration

Input: $T \in \mathbb{R}^{n \times n \times n}$.
Initialize: $\widetilde{T} := T$.
For $i = 1, 2, \ldots, n$:

1. Pick $\vec{x}^{(0)} \in \mathbb{S}^{n-1}$ unif. at random.
2. Run tensor power iteration with $\widetilde{T}$ starting from $\vec{x}^{(0)}$ for N iterations.
3. Set $\hat{v}_i := \vec{x}^{(N)} / \|\vec{x}^{(N)}\|$ and $\hat{\lambda}_i := f_{\widetilde{T}}(\hat{v}_i)$.
4. Replace $\widetilde{T} := \widetilde{T} - \hat{\lambda}_i \, \hat{v}_i \otimes \hat{v}_i \otimes \hat{v}_i$.

Output: $\{(\hat{v}_i, \hat{\lambda}_i) : \; i \in [n]\}$.

Power iteration

- Allows you to find individual eigenvalue / eigenvector pairs
- Matrix: linearly quickly $O\left(\log \frac{1}{||\mathcal{E}||}\right)$
- Tensor: quadratically quickly $O\left(\log \log \frac{1}{||\mathcal{E}||}\right)$
- Both require gap between largest and second-largest θ_i

Alternative: Direct Minimization

$$\left\| T - \sum_t \theta_t \phi_t \otimes \phi_t \otimes \phi_t \right\|_F^2 \tag{9}$$

- Use gradient descent to directly optimize parameters
- Wins over "standard" approaches because fewer observations
- Disliked by theory folks

Spectral Methods

- If you only care about high-level patterns
- You can often get that from statistical summaries
- **Ignore the data!**
- These approaches often have nice runtimes