Discovery of Latent Factors in High-dimensional Data via Spectral Methods

Furong Huang

University of Maryland

Workshop on Quantum Machine Learning

Machine Learning - Excitements

Success of Supervised Learning


Image classification


Speech recognition


Text processing

Machine Learning - Excitements

Success of Supervised Learning


Image classification


Speech recognition


Text processing

Key to Success

- Deep composition of nonlinear units
- Enormous labeled data
- Computation power growth

Machine Learning - Modern Challenges

Automated discovery of features and categories?

Filter bank learning

Feature extraction

Embeddings, Topics

Machine Learning - Modern Challenges

Automated discovery of features and categories?

Real AI requires **Unsupervised Learning**

Filter bank learning

Feature extraction

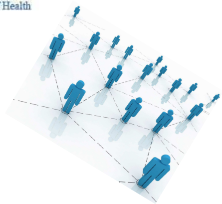
Embeddings, Topics

- Summarize key features in data
  - State-of-the-art: Humans are better than machines
  - Goal: Intelligent machines that summarize key features in data
- Interpretable modeling and learning of the data
  - Theoretically guaranteed learning
  - Extracted features are interpretable

Unsupervised Learning with Big Data

Curse of Dimensionality

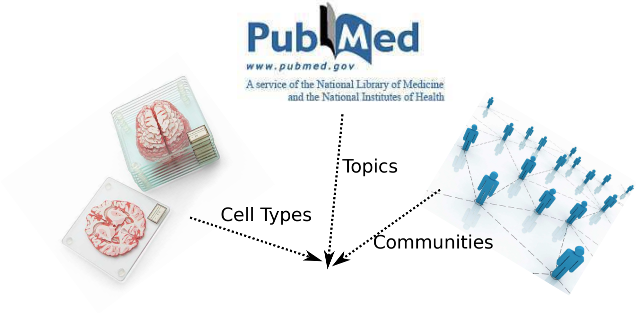
- More information $\rightarrow$ more unknowns/variables $\rightarrow$ **challenging model learning**



Unsupervised Learning with Big Data

Information Extraction

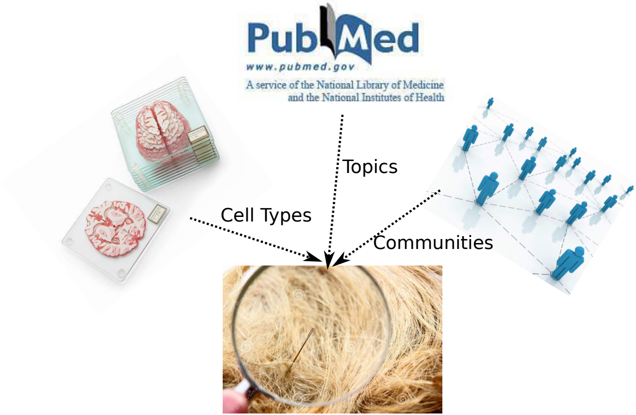
- High dimension **observation** vs Low dimension **representation**



Unsupervised Learning with Big Data

Information Extraction

- High dimension **observation** vs Low dimension **representation**

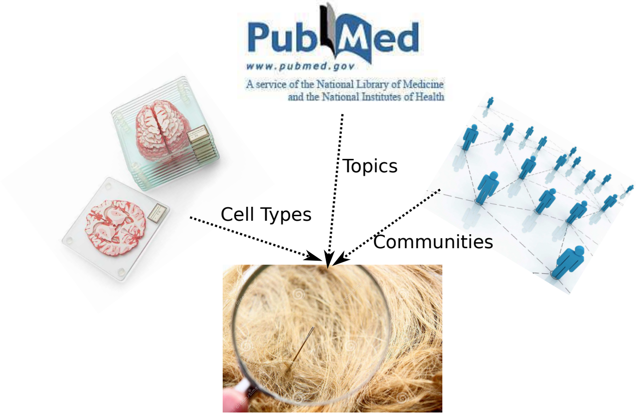


Finding Needle In the Haystack Is Challenging

Unsupervised Learning with Big Data

Information Extraction

- High dimension **observation** vs Low dimension **representation**



My Solution: A Unified Tensor Decomposition Framework

App 1: Automated Categorization of Documents


SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor — I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five

am's second-leading receiver.

On Jan. 10, 2013, a female student at Florida State spotted the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

In the 21 months since, Florida State officials have said little about how they handled the case, which is no

investigated by the federal Depart

Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.

"Can you share any details on the requesting source?" David Perry, the university's police chief, asked the Tallahassee police. Several hours later, Mr.

Document modeling

- Observed: words in document corpus: search logs, emails etc
- Hidden: (mixed) topics: personal interests, professional area etc

App 1: Automated Categorization of Documents

SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor — I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five

am's second-leading receiver.

On Jan. 10, 2013, a female student at Florida State spotted the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

In the 21 months since, Florida State officials have said little about how they handled the case, which is no

investigated by the federal Depart

Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

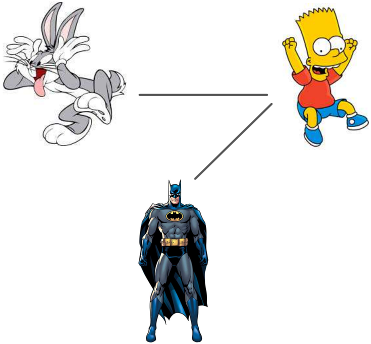
Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.

"Can you share any details on the requesting source?" David Perry, the university's police chief, asked the Tallahassee police. Several hours later, Mr.

Document modeling

- **Observed:** words in document corpus: search logs, emails etc
- **Hidden:** (mixed) topics: personal interests, professional area etc

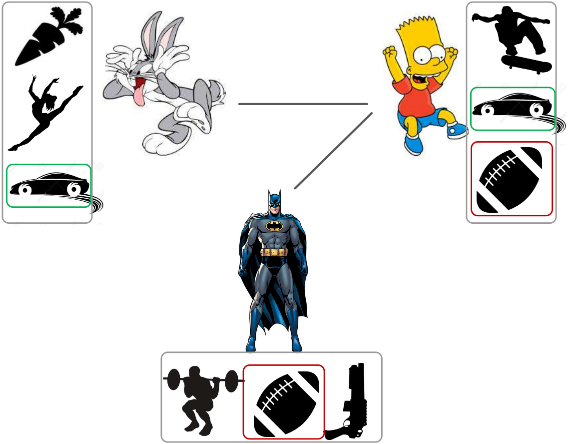
App 2: Community Extraction From Connectivity Graph



Social Networks

- Observed: network of social ties: friendships, transactions etc
- Hidden: (mixed) groups/communities of social actors

App 2: Community Extraction From Connectivity Graph

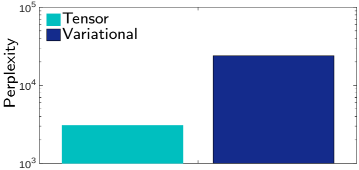


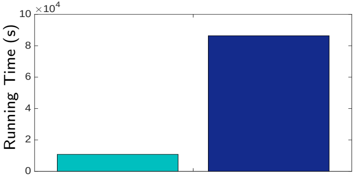
Social Networks

- Observed: network of social ties: friendships, transactions etc
- Hidden: (mixed) groups/communities of social actors

Tensor Methods Compared with Variational Inference

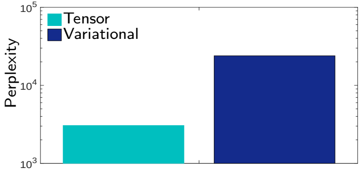
Learning Topics from PubMed on Spark: 8 million docs

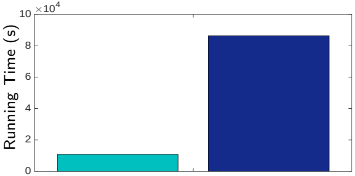




Tensor Methods Compared with Variational Inference

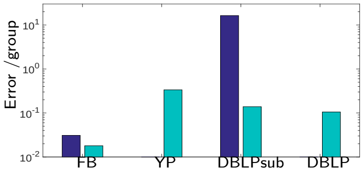
Learning Topics from PubMed on Spark: 8 million docs

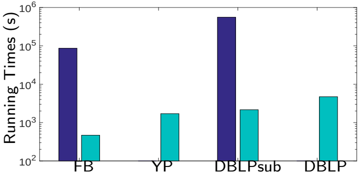




Learning Communities from Graph Connectivity

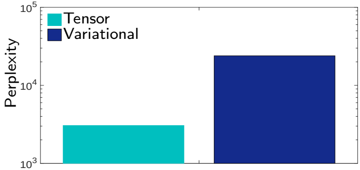
Facebook: $n \sim 20k$ Yelp: $n \sim 40k$ DBLPsub: $n \sim 0.1m$ DBLP: $n \sim 1m$

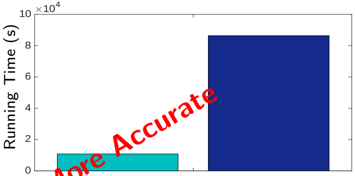




Tensor Methods Compared with Variational Inference

Learning Topics from PubMed on Spark: 8 million docs

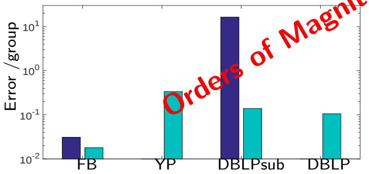


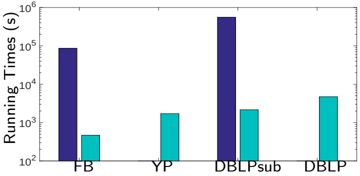


Orders of Magnitude Faster & More Accurate

Learning Communities from Graph Connectivity

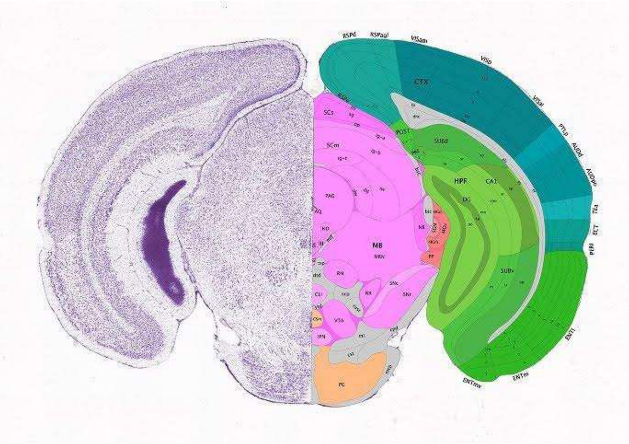
Facebook: $n \sim 20k$ Yelp: $n \sim 40k$ DBLPsub: $n \sim 0.1m$ DBLP: $n \sim 1m$





"Online Tensor Methods for Learning Latent Variable Models", F. Huang, U. Niranjan, M. Hakeem, A. Anandkumar, JMLR14.
"Tensor Methods on Apache Spark", F. Huang, A. Anandkumar, Oct. 2015.

App 3: Cataloging Neuronal Cell Types In the Brain

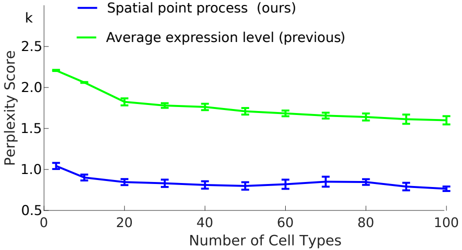


Neuroscience

- Observed: cellular-resolution brain slices
- Hidden: neuronal cell types

App 3: Cataloging Neuronal Cell Types In the Brain

- Our method vs Average expression level [Grange 14']

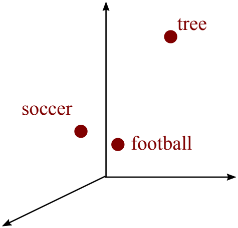


Recovered known cell types

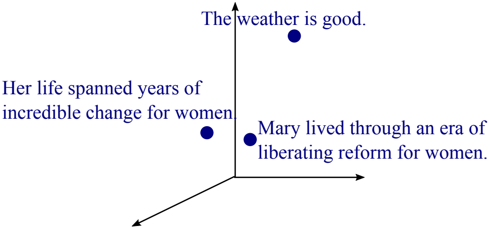
1. *Interneurons*
2. *S1Pyramidal*
3. *Astrocytes*
4. *Ependymal*
5. *Microglia*
6. *Endothelial*
7. *Mural*
8. *Oligodendrocytes*

"Discovering Neuronal Cell Types and Their Gene Expression Profiles Using a Spatial Point Process Mixture Model", F. Huang, A. Anandkumar, C. Borgs, J. Chayes, E. Fraenkel, M. Hawrylycz, E. Lein, A. Ingrosso, S. Turaga, NIPS 2015 BigNeuro workshop.

App 4: Word Sequence Embedding Extraction



Word Embedding

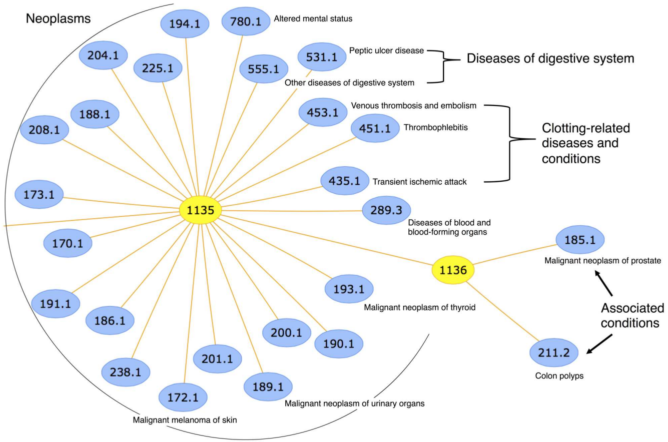


Word Sequence Embedding

"Convolutional Dictionary Learning through Tensor Factorization", by F. Huang, A. Anandkumar, In Proceedings of JMLR 2015.

App 5: Human Disease Hierarchy Discovery

CMS: 1.6 million patients, 168 million diagnostic events, 11 k diseases.



- Observed: co-occurrence of diseases on patients
- Hidden: disease similarity/hierarchy

" Scalable Latent TreeModel and its Application to Health Analytics " by F. Huang, N. U.Niranjan, I. Perros, R. Chen, J. Sun, A. Anandkumar, NIPS 2015 MLHC workshop.

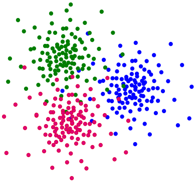
$\boxed{\text{Involve discovering the hidden and compact structure}}$

$\boxed{\text{that is embedded in the high-dimensional complex observed data}}$

How to model hidden effects?

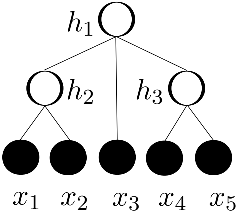
Basic Approach: mixtures/clusters

- Hidden variable h is categorical.



Advanced: Probabilistic models

- Hidden variable h has more general distributions.
- Can model mixed memberships.

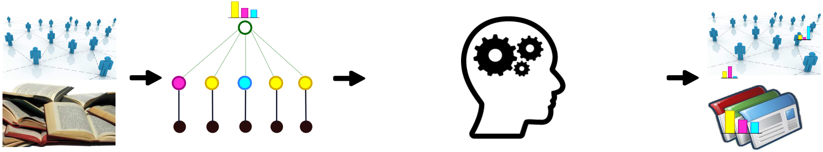


This talk: basic mixture model and some advanced models (topic model)

Challenges in Learning

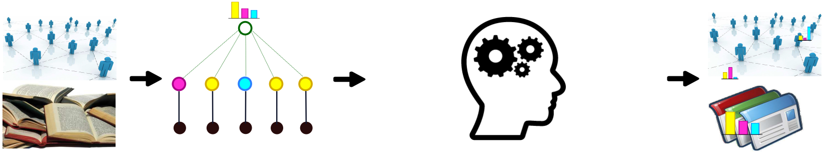
Basic goal in all mentioned applications

Discover hidden structure in data: unsupervised learning.



Unlabeled data Latent variable model Learning Algorithm Inference

Challenges in Learning – find hidden structure in data

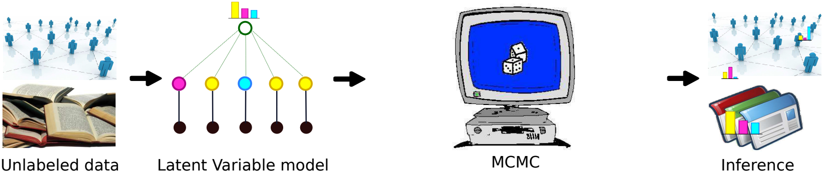


Unlabeled data | Latent variable model | Learning Algorithm | Inference

Challenge: Conditions for Identifiability

- Whether can model be identified given infinite computation and data?
- Are there tractable algorithms under identifiability?

Challenges in Learning – find hidden structure in data



Challenge: Conditions for Identifiability

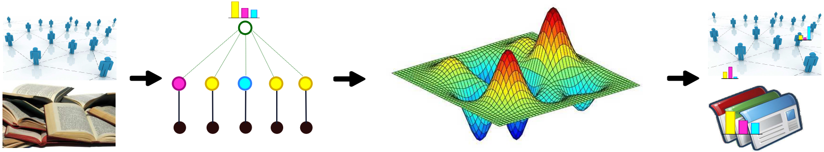
- Whether can model be identified given infinite computation and data?
- Are there tractable algorithms under identifiability?

Challenge: Efficient Learning of Latent Variable Models

- MCMC: random sampling, slow

  Exponential mixing time

Challenges in Learning – find hidden structure in data



Unlabeled data | Latent variable model | Likelihood Methods | Inference

Challenge: Conditions for Identifiability

- Whether can model be identified given infinite computation and data?
- Are there tractable algorithms under identifiability?

Challenge: Efficient Learning of Latent Variable Models

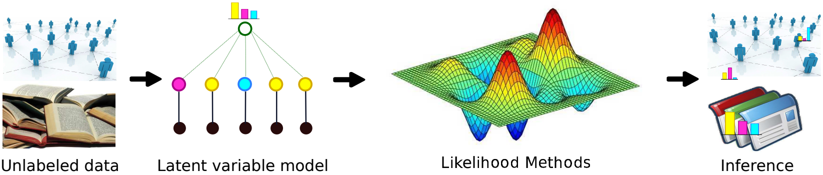
- MCMC: random sampling, slow

  Exponential mixing time

- Likelihood: non-convex, not scalable

  Exponential critical points

Challenges in Learning – find hidden structure in data



Challenge: Conditions for Identifiability

- Whether can model be identified given infinite computation and data?
- Are there tractable algorithms under identifiability?

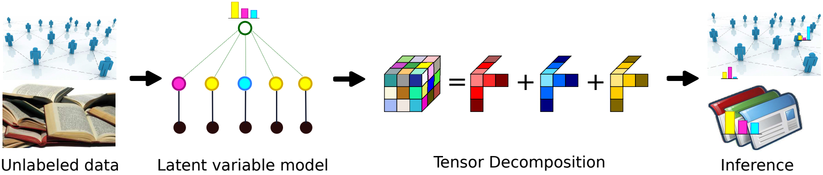
Challenge: Efficient Learning of Latent Variable Models

- MCMC: random sampling, slow

  Exponential mixing time
- Likelihood: non-convex, not scalable

  Exponential critical points
- Efficient computational and sample complexities?

Challenges in Learning – find hidden structure in data



Challenge: Conditions for Identifiability

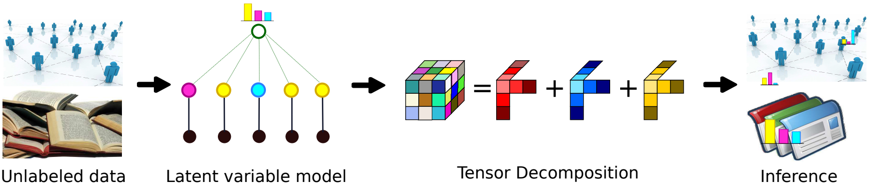
- Whether can model be identified given infinite computation and data?
- Are there tractable algorithms under identifiability?

Challenge: Efficient Learning of Latent Variable Models

- MCMC: random sampling, slow
  - Exponential mixing time
- Likelihood: non-convex, not scalable
  - Exponential critical points
- Efficient computational and sample complexities?

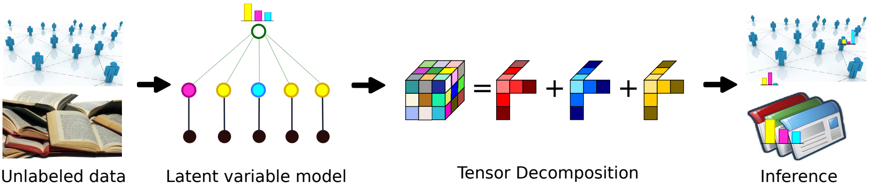
Guaranteed and efficient learning through spectral methods

Unsupervised Learning via Probabilistic Models



$\boxed{\text{tensor decomposition} \rightarrow \text{correct model}}$

Unsupervised Learning via Probabilistic Models



$\boxed{\text{tensor decomposition} \rightarrow \text{correct model}}$

Contributions

- Guaranteed online algorithm with global convergence guarantee
- Highly scalable, highly parallel, dimensionality reduction
- Tensor library on CPU/GPU/Spark
- Interdisciplinary applications
- Extension to model with group invariance

Outline

Method-of-Moments At A Glance

1. Determine function of model parameters $\boldsymbol{\theta}$ estimatable from observable data:
   - **Moments**

$$\mathbb{E}_{\boldsymbol{\theta}}[f(\boldsymbol{X})]$$

2. Form estimates of moments using data (iid samples $\{\boldsymbol{x}_i\}_{i=1}^n$):
   - **Empirical Moments**

$$\widehat{\mathbb{E}}[f(\boldsymbol{X})]$$

3. Solve the approximate equations for parameters $\boldsymbol{\theta}$:
   - **Moment matching**

$$\mathbb{E}_{\boldsymbol{\theta}}[f(\boldsymbol{X})] \overset{n\to\infty}{=} \widehat{\mathbb{E}}[f(\boldsymbol{X})]$$

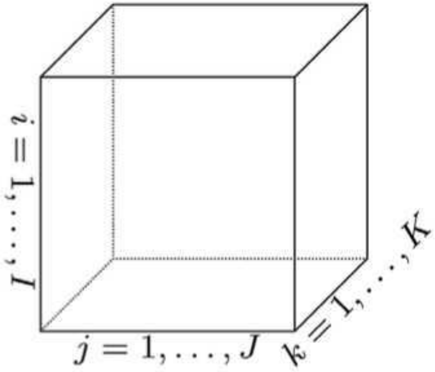
Toy Example

How to estimate Gaussian variable, i.e., $(\boldsymbol{\mu},\boldsymbol{\Sigma})$,
given iid samples $\{\boldsymbol{x}_i\}_{i=1}^n \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}^2)$?

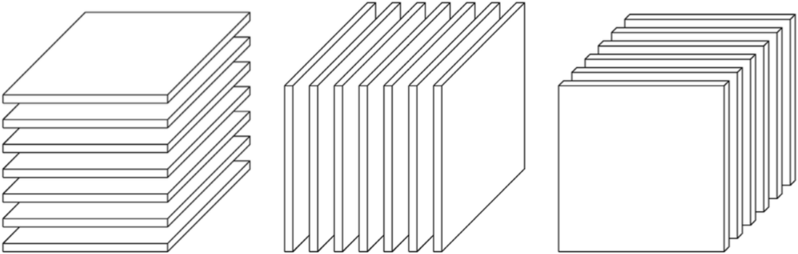
What is a tensor?

Multi-dimensional Array

- Tensor - Higher order matrix
- The number of dimensions is called tensor order.

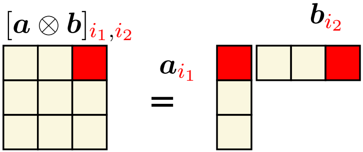


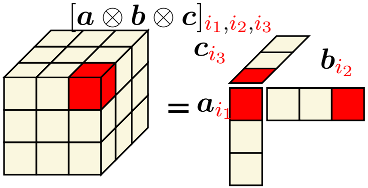
Slices



- Horizontal slices
- Lateral slices
- Frontal slices

Tensor Product





- $[\boldsymbol{a} \otimes \boldsymbol{b}]_{i_1,i_2} = \boldsymbol{a}_{i_1} \boldsymbol{b}_{i_2}$
- Rank-1 matrix

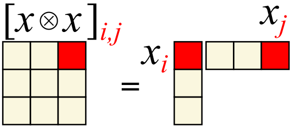
- $[\boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c}]_{i_1,i_2,i_3} = \boldsymbol{a}_{i_1} \boldsymbol{b}_{i_2} \boldsymbol{c}_{i_3}$
- Rank-1 tensor

Tensors in Method of Moments

Matrix: Pair-wise relationship

- Signal or data observed $\boldsymbol{x} \in \mathbb{R}^d$
- Rank 1 matrix: $[\boldsymbol{x} \otimes \boldsymbol{x}]_{i,j} = \boldsymbol{x}_i \boldsymbol{x}_j$
- Aggregated pair-wise relationship

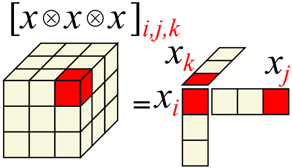
$$\boldsymbol{M}_2 = \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x}]$$



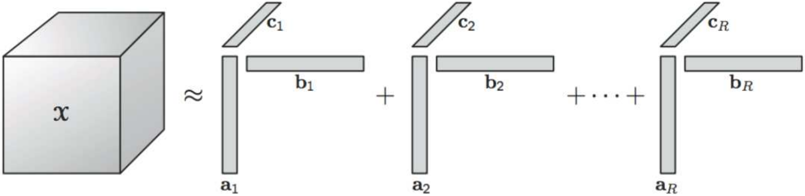
Tensor: Triple-wise relationship or higher

- Signal or data observed $\boldsymbol{x} \in \mathbb{R}^d$
- Rank 1 tensor:
  $[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}]_{i,j,k} = \boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{x}_k$
- Aggregated triple-wise relationship

$$\boldsymbol{\mathcal{M}}_3 = \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] = \mathbb{E}[\boldsymbol{x} \otimes^3]$$



CP decomposition



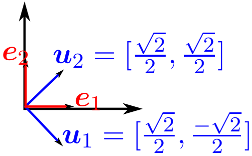
- $\boldsymbol{\mathcal{X}} = \sum_{h=1}^{R} \boldsymbol{a}_h \otimes \boldsymbol{b}_h \otimes \boldsymbol{c}_h$
- Summation of rank-1 tensors

Why are tensors powerful?

Matrix Orthogonal Decomposition

- Not unique without eigenvalue gap

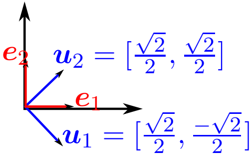
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{e}_2\boldsymbol{e}_2^\top = \boldsymbol{u}_1\boldsymbol{u}_1^\top + \boldsymbol{u}_2\boldsymbol{u}_2^\top$$



Why are tensors powerful?

Matrix Orthogonal Decomposition

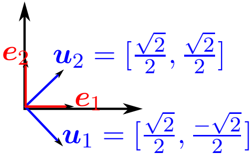
- Not unique without eigenvalue gap
  $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{e}_2\boldsymbol{e}_2^\top = \boldsymbol{u}_1\boldsymbol{u}_1^\top + \boldsymbol{u}_2\boldsymbol{u}_2^\top$$
- Unique with eigenvalue gap



Why are tensors powerful?

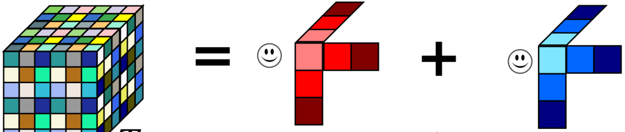
Matrix Orthogonal Decomposition

- Not unique without eigenvalue gap
  $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$
- Unique with eigenvalue gap



Tensor Orthogonal Decomposition (Harshman, 1970)

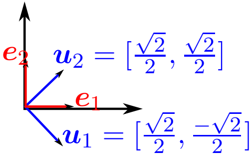
- Unique: eigenvalue gap not needed



$$Tensor = u_1 \otimes u_1 \otimes u_1 + u_2 \otimes u_2 \otimes u_2$$

Why are tensors powerful?

Matrix Orthogonal Decomposition



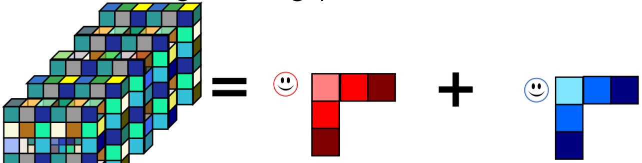
- Not unique without eigenvalue gap

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = e_1 e_1^\top + e_2 e_2^\top = u_1 u_1^\top + u_2 u_2^\top$$

- Unique with eigenvalue gap

Tensor Orthogonal Decomposition (Harshman, 1970)

- Unique: eigenvalue gap not needed
- Slice of tensor has eigenvalue gap



$$Slice_i = u_1(i)\, u_1 \otimes u_1 + u_2(i)\, u_2 \otimes u_2$$

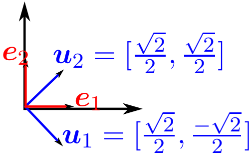
Why are tensors powerful?

Matrix Orthogonal Decomposition

- Not unique without eigenvalue gap

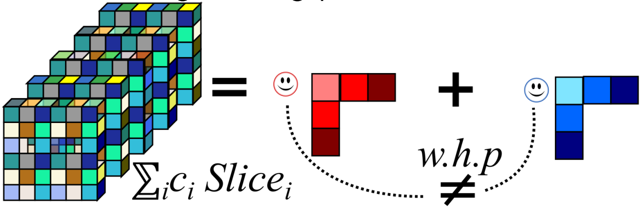
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{e}_2\boldsymbol{e}_2^\top = \boldsymbol{u}_1\boldsymbol{u}_1^\top + \boldsymbol{u}_2\boldsymbol{u}_2^\top$$

- Unique with eigenvalue gap



Tensor Orthogonal Decomposition (Harshman, 1970)

- Unique: eigenvalue gap not needed
- Slice of tensor has eigenvalue gap



Outline

Outline

Probabilistic Topic Models - LDA

Bag of words

- Infer topics of documents
- Learn hidden process drives the obs.

Generative model

- Topic proportion $\sim$ Dir(α) for a doc
- Draw a topic, then a word for a token

Probabilistic Topic Models - LDA

Bag of words

- Infer topics of documents
- Learn hidden process drives the obs.

Generative model

- Topic proportion $\sim$ Dir(α) for a doc
- Draw a topic, then a word for a token

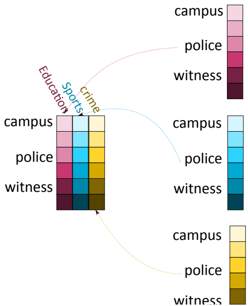
Probabilistic Topic Models - LDA

Bag of words

- Infer topics of documents
- Learn hidden process drives the obs.

Generative model

- Topic proportion $\sim \mathrm{Dir}(\alpha)$ for a doc
- Draw a topic, then a word for a token



SECTIONS HOME SEARCH The New York Times

COLLEGE FOOTBALL

At Florida State, Football Clouds Justice

By MIKE McINTIRE and WALT BOGDANICH OCT. 10, 2014

Now, an examination by The New York Times of police and court records, along with interviews with crime witnesses, has found that, far from an aberration, the treatment of the Winston complaint was in keeping with the way the police on numerous occasions have soft-pedaled allegations of wrongdoing by Seminoles football players. From criminal mischief and motor-vehicle theft to domestic violence, arrests have been avoided, investigations have stalled and players have escaped serious consequences.

In a community whose self-image and economic well-being are so tightly bound to the fortunes of the nation's top-ranked college football team, law enforcement officers are finely attuned to a suspect's football connections. Those ties are cited repeatedly in police reports examined by The Times. What's more, dozens of officers work second jobs directing traffic and providing security at home football games, and many express their devotion to the Seminoles on social media.

TMZ, the gossip website, also requested the police report and later asked the school's deputy police chief, Jim L. Russell, if the campus police had interviewed Mr. Winston about the rape report. Mr. Russell responded by saying his officers were not investigating the case, omitting any reference to the city police, even though the campus police knew of their involvement. "Thank you for contacting me regarding this rumor — I am glad I can dispel that one!" Mr. Russell told TMZ in an email. The university said Mr. Russell was unaware of any other police investigation at the time of the inquiry. Soon after, the Tallahassee police belatedly sent their files to the news media and to the prosecutor, William N. Meggs. By then critical evidence had been lost and Mr. Meggs, who criticized the police's handling of the case, declined to

lson after the Seminoles' first game; five

am's second-leading receiver.

On Jan. 10, 2013, a female student at Florida State spotted the man she believed had raped her the previous month. After learning his name, Jameis Winston, she reported him to the Tallahassee police.

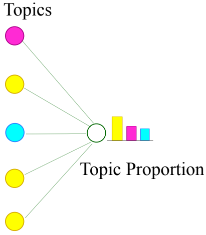
In the 21 months since, Florida State officials have said little about how they handled the case, which is no

investigated by the federal Depart

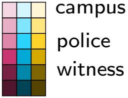
Most recently, university officials suspended Mr. Winston for one game after he stood in a public place on campus and, playing off a running Internet gag, shouted a crude reference to a sex act. In a news conference afterward, his coach, Jimbo Fisher, said, "Our hope and belief is Jameis will learn from this and use better judgment and language and decision-making."

As The Times reported last April, the Tallahassee police also failed to aggressively investigate the rape accusation. It did not become public until November, when a Tampa reporter, Matt Baker, acting on a tip, sought records of the police investigation.

Upon learning of Mr. Baker's inquiry, Florida State, having shown little curiosity about the rape accusation, suddenly took a keen interest in the journalist seeking to report it, according to emails obtained by The Times.



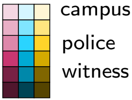
Goal



- Topic-word matrix $\boxed{\mathbb{P}[\text{word} = e_i | \text{topic} = j]}$

Moments Matching

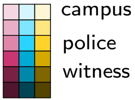
Goal: Linearly independent topic-word table



$$\boxed{\mathbb{E}[\text{word}|\text{topic} = j] = \sum_i \mathbb{P}[\text{word} = e_i|\text{topic} = j] e_i = \text{column } j}$$

Moments Matching

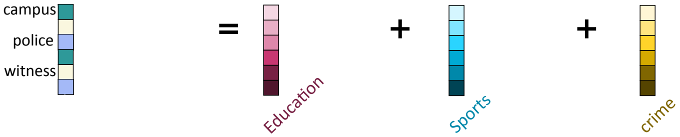
Goal: Linearly independent topic-word table



$$\boxed{\mathbb{E}[\text{word}|\text{topic}=j]=\sum_i \mathbb{P}[\text{word}=e_i|\text{topic}=j]e_i=\text{column } j}$$

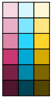
M_1: Occurrence Frequency of Words

$$\mathbb{E}[\text{word}]=\sum_j \mathbb{E}[\text{word}|\text{topic}=j]\mathbb{P}[\text{topic}=j]$$



Moments Matching

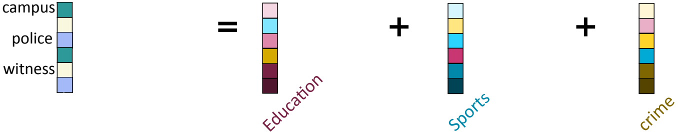
Goal: Linearly independent topic-word table



$$\boxed{\mathbb{E}[\text{word}|\text{topic} = j] = \sum_i \mathbb{P}[\text{word} = e_i|\text{topic} = j]e_i = \text{column } j}$$

M_1: Occurrence Frequency of Words

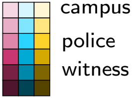
$$\mathbb{E}[\text{word}] = \sum_j \mathbb{E}[\text{word}|\text{topic} = j]\mathbb{P}[\text{topic} = j]$$



No unique decomposition of vectors

Moments Matching

Goal: Linearly independent topic-word table



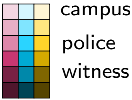
$$\boxed{\mathbb{E}[\text{word}|\text{topic}=j] = \sum_i \mathbb{P}[\text{word}=e_i|\text{topic}=j]e_i = \text{column } j}$$

M_2: Modified Co-occurrence Frequency of Word Pairs

$$\mathbb{E}[\text{word}_1 \otimes \text{word}_2] = \sum_{j,j'} \mathbb{E}[\text{word}_1|\text{topic}_1=j] \otimes \mathbb{E}[\text{word}_2|\text{topic}_2=j']\mathbb{P}[\text{topic}_1=j,\text{topic}_2=j']$$

Moments Matching

Goal: Linearly independent topic-word table



$$\boxed{\mathbb{E}[\text{word}|\text{topic}=j] = \sum_i \mathbb{P}[\text{word}=e_i|\text{topic}=j]e_i = \text{column } j}$$

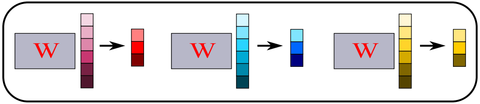
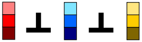
M_2: Modified Co-occurrence Frequency of Word Pairs

$$\mathbb{E}[\text{word}_1 \otimes \text{word}_2] = \sum_{j,j'} \mathbb{E}[\text{word}_1|\text{topic}_1 = j] \otimes \mathbb{E}[\text{word}_2|\text{topic}_2 = j']\mathbb{P}[\text{topic}_1 = j, \text{topic}_2 = j']$$

Matrix decomposition recovers subspace, not actual model

Moments Matching

Goal: Linearly independent topic-word table

Find a W  such that 

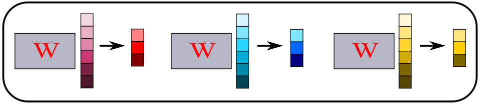
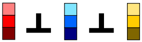
M_2: Modified Co-occurrence Frequency of Word Pairs

$$\mathbb{E}[\text{word}_1 \otimes \text{word}_2] = \sum_{j,j'} \mathbb{E}[\text{word}_1 | \text{topic}_1 = j] \otimes \mathbb{E}[\text{word}_2 | \text{topic}_2 = j'] \mathbb{P}[\text{topic}_1 = j, \text{topic}_2 = j']$$

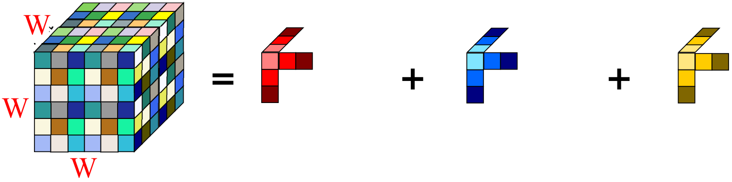
$\boxed{\text{Many such } W\text{'s, find one, project data with } W}$

Moments Matching

Goal: Linearly independent topic-word table

Know a W  such that 

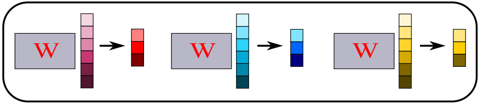
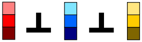
M_3: Modified Co-occurrence Frequency of Word Triplets



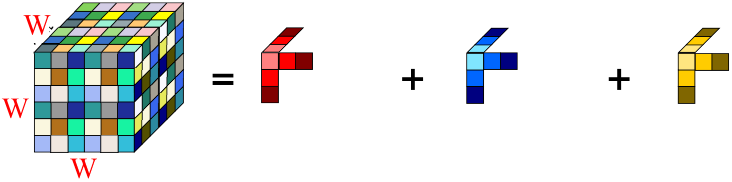
Unique orthogonal tensor decomposition, project result with $W^\dagger$

Moments Matching

Goal: Linearly independent topic-word table

Know a W  such that 

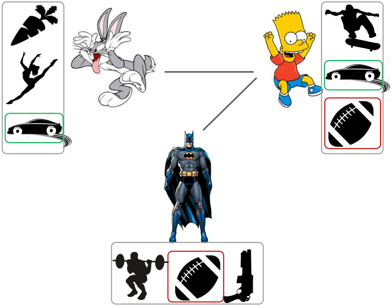
M_3: Modified Co-occurrence Frequency of Word Triplets



Tensor decomposition uniquely discovers the correct model

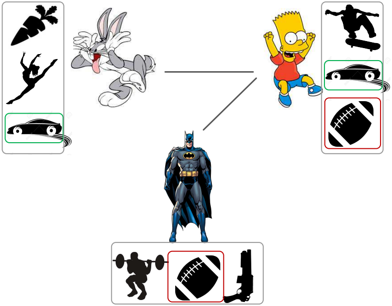
Mixed Membership Community Models

Mixed memberships

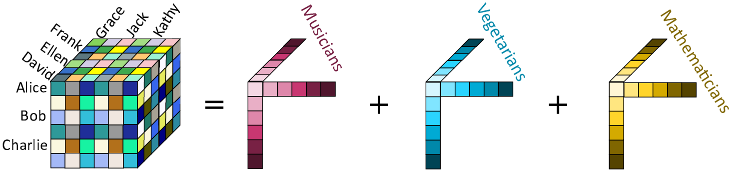


Mixed Membership Community Models

Mixed memberships

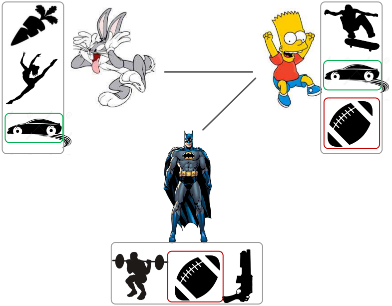


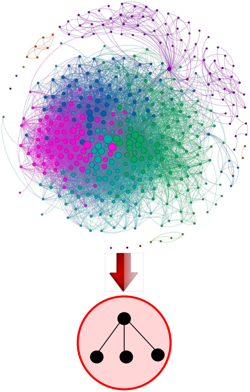
What ensures guaranteed learning?



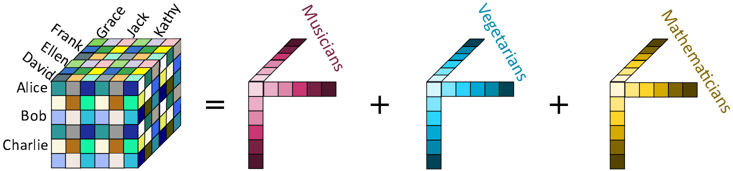
Mixed Membership Community Models

Mixed memberships





What ensures guaranteed learning?



Outline

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i,$ where $\|a_i\| = 1, a_i^\top a_j = 0$

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i,$ where $\|a_i\| = 1, a_i^\top a_j = 0$

Objective?

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i,$ where $\|a_i\| = 1, a_i^\top a_j = 0$

Objective?

- Objective $\min_{\forall i, \|u_i\|^2=1} \quad \sum_{i \neq j} T(u_i, u_i, u_j, u_j)$ **Non-convex!**

Theorem: The proposed objective function has equivalent local optima.

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

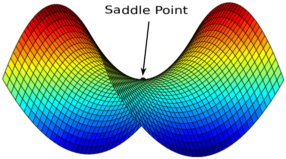
- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i,$ where $\|a_i\| = 1, a_i^\top a_j = 0$

Objective?

- Objective $\min_{\forall i, \|u_i\|^2=1} \quad \sum_{i \neq j} T(u_i, u_i, u_j, u_j)$ **Non-convex!**

Theorem: The proposed objective function has equivalent local optima.

Will SGD work?



Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

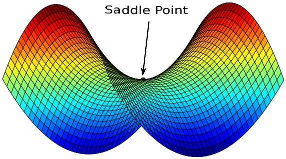
- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i$, where $\|a_i\| = 1, a_i^\top a_j = 0$

Objective?

- Objective $\min_{\forall i, \|u_i\|^2=1} \sum_{i \neq j} T(u_i, u_i, u_j, u_j)$ **Non-convex!**

Theorem: The proposed objective function has equivalent local optima.

Will SGD work?



Theorem: For smooth, twice-diff fn. with non-degenerate saddle points, noisy SGD converges to a local optimum in polynomial steps.

Guaranteed Online Tensor Decomposition

Model is uniquely identifiable! How to identify?

Online Tensor Decomposition

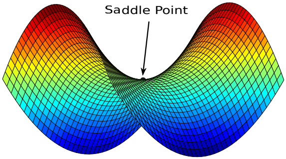
- Tensor $T = \sum_i a_i \otimes a_i \otimes a_i \otimes a_i$, where $\|a_i\| = 1, a_i^\top a_j = 0$

Objective?

- Objective $\min_{\forall i, \|u_i\|^2=1} \sum_{i \neq j} T(u_i, u_i, u_j, u_j)$ **Non-convex!**

Theorem: The proposed objective function has equivalent local optima.

Will SGD work?

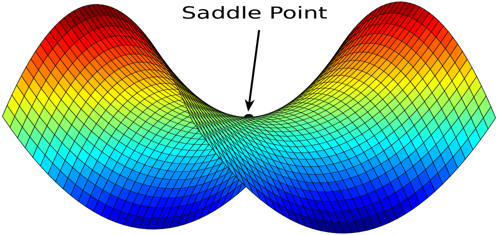


Theorem: For smooth, twice-diff fn. with non-degenerate saddle points, noisy SGD converges to a local optimum in polynomial steps.

Global Convergence Guarantee For Online Tensor Decomposition

Why could we escape from saddle points?

Stochastic Gradient Descent with Noise

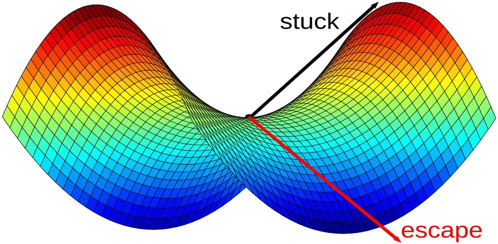


- Saddle point has 0 gradient

"Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition",by R. Ge, F. Huang, C. Jin, Y. Yuan, COLT 2015.

Why could we escape from saddle points?

Stochastic Gradient Descent with Noise

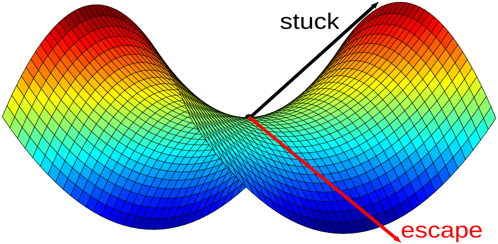


- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue

"Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition",by R. Ge, F. Huang, C. Jin, Y. Yuan, COLT 2015.

Why could we escape from saddle points?

Stochastic Gradient Descent with Noise

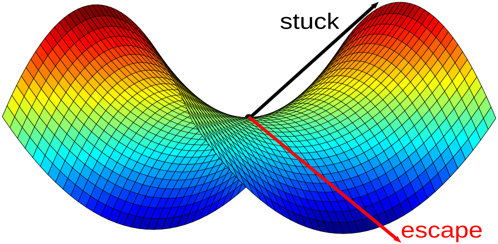


- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue
- Negative eigenvalue: direction of escape

"Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition",by R. Ge, F. Huang, C. Jin, Y. Yuan, COLT 2015.

Why could we escape from saddle points?

Stochastic Gradient Descent with Noise



- Saddle point has 0 gradient
- Non-degenerate saddle: Hessian has $\pm$ eigenvalue
- Negative eigenvalue: direction of escape

Noise could help!

"Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition",by R. Ge, F. Huang, C. Jin, Y. Yuan, COLT 2015.

Outline

First PCA

PCA problem

- Sample $S = \{\boldsymbol{x}_i\}_{i=1}^m$, where $\boldsymbol{x}_i \in \mathbb{R}^d$
- Q: Identifies the direction of the largest variance in the data?

Problem Formulation

Solving

$$\max_{\boldsymbol{u} \in \mathbb{R}^d, \|\boldsymbol{u}\|_2 = 1} \boldsymbol{u}^\top \boldsymbol{A} \boldsymbol{u},$$

where covariance matrix $\boldsymbol{A} = \frac{1}{m} \sum_{i=1}^m \boldsymbol{x}_i \boldsymbol{x}_i^\top$

Problem Regime

- Assume $0 \preceq A \preceq I$ and s-sparse (i.e., nnz(each row or column) $\leq s$)

Classical Algorithm

Spectral Gap

- Spectral gap $\Delta = \lambda_1 - \lambda_2$
  - ordered eigenvalues $1 \geq \lambda_1 \geq \ldots \lambda_d \geq 0$
  - and corresponding eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d$.

Methods under Warm Start

- Warm start: Initialization $\boldsymbol{v}_0$ such that $|< \boldsymbol{v}_0, \boldsymbol{u}_1 >| > \phi > 0$
- Iteration methods achieve ϵ precision: $< \boldsymbol{v}_k, \boldsymbol{u}_1 > \geq 1 - \epsilon$
  - Power method $\frac{\boldsymbol{A}^k \boldsymbol{v}_0}{\|\boldsymbol{A}^k \boldsymbol{v}_0\|}$ takes $O(\frac{sd}{\Delta} \log(\frac{1}{\phi\epsilon}))$
  - Lanczos method or accelerated power method takes $O(\frac{sd}{\sqrt{\Delta}} \log(\frac{1}{\phi\epsilon}))$
    - Replacing the monomial $\boldsymbol{A}^k$ by its Chebyshev polynomial approximation

---

Question: Speedup from $O(d)$ to $\mathrm{poly}(\log d)$?

Quantum Speedup

Motivation

Quantum effects can achieve significant speedup.

Examples

- Shor's algorithm
  - exponential speed-up for factoring integers
- Grover's algorithm
  - quadratic speed-up for searching in unstructured database
- (Harrow, Hassidim, Lloyd '09) & (Childs, Kothari, Somma '17)
  - $\Omega(d) \to \mathrm{poly}(\log d)$ for solving d-dimensional linear equation systems.
  - weaker output requirement
    - a quantum state whose vector representation is roughly the solution to the linear equation system.

Quantum Leading PCA

Input model

- Quantum oracle which generates a quantum state whose vector representation is $\boldsymbol{v}_0$ and $\boldsymbol{A}$.

Output model

- A quantum state whose vector representation is $\boldsymbol{v}_k$

Main Result

Under warm start $| < \boldsymbol{v}_0, \boldsymbol{u}_1 > | = \phi > 0$, there is a quantum algorithm which prepares a quantum state with vector representation $\boldsymbol{v}_k$ such that $< \boldsymbol{v}_k, \boldsymbol{u}_1 > \geq 1 - \epsilon$ with probability at least 2/3

- using $O(s \log(s/\phi\epsilon)/\phi\sqrt{\Delta})$ queries to quantum oracle $U_{A,s}$, $U_{A,e}$
- $O(1/\phi)$ queries to $U_{\boldsymbol{v}_0}$

w. $O(s(\log d \log(\frac{s}{\phi\epsilon}) + \log^{3.5}(\frac{s}{\phi\epsilon}))/\phi\sqrt{\Delta})$ 2-qubit quantum gates in total.

Joint work with Tongyang Li and Xiaodi Wu.

Intuition for Speedup

Chebyshev polynomials can be significantly accelerated in quantum computation

- Matrix power $\boldsymbol{A}^k\boldsymbol{b}$ is the key
  - Quantum-walk
    - ⋆ effectively constructs a degree-m Chebyshev polynomial of $\boldsymbol{A}/s$.
  - Quantum primitive: the linear combination of unitaries (LCU)
    - ⋆ effectively linearly combines these Chebyshev polynomials to derive the desired approximation polynomial.

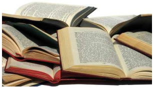
Quantum Computation for Linear Algebraic Problems

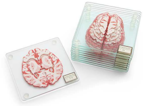
Outline

Summary

Spectral methods reveal hidden structure

- Text/Image processing
- Social networks
- Neuroscience, healthcare ...

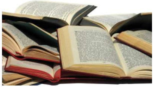


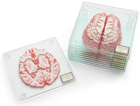


Summary

Spectral methods reveal hidden structure

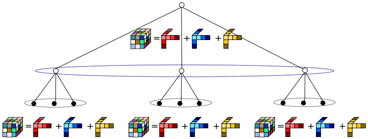
- Text/Image processing
- Social networks
- Neuroscience, healthcare ...

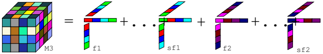


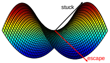


Versatile for latent variable models

- Flat model → hierarchical model
- Sparse coding → convolutional model
- Efficient, convergence guarantee







Thank You

furongh@cs.umd.edu