Tensor Methods for large-scale Machine Learning

Anima Anandkumar

U.C. Irvine
Learning with Big Data
Data vs. Information
Data vs. Information

This magazine should have some amusing ads soon.

I love messing with data.
Data vs. Information

- Missing observations, gross corruptions, outliers.
Data vs. Information

- Missing observations, gross corruptions, outliers.
- High dimensional regime: as data grows, more variables!
Data vs. Information

- Missing observations, gross corruptions, outliers.
- High dimensional regime: as data grows, more variables!

Data deluge an information desert!
Learning in High Dimensional Regime

- Useful information: low-dimensional structures.
- Learning with big data: ill-posed problem.
Learning in High Dimensional Regime

- Useful information: low-dimensional structures.
- Learning with big data: ill-posed problem.

Learning is finding needle in a haystack
Learning in High Dimensional Regime

- Useful information: low-dimensional structures.
- Learning with big data: ill-posed problem.

Learning is finding needle in a haystack

Learning with big data: computationally challenging!

Principled approaches for finding low dimensional structures?
How to model information structures?

Latent variable models

- Incorporate hidden or latent variables.
- Information structures: Relationships between latent variables and observed data.
How to model information structures?

Latent variable models

- Incorporate hidden or latent variables.
- Information structures: Relationships between latent variables and observed data.

Basic Approach: mixtures/clusters

- Hidden variable is categorical.
How to model information structures?

Latent variable models
- Incorporate hidden or latent variables.
- Information structures: Relationships between latent variables and observed data.

Basic Approach: mixtures/clusters
- Hidden variable is categorical.

Advanced: Probabilistic models
- Hidden variables have more general distributions.
- Can model mixed membership/hierarchical groups.
Latent Variable Models (LVMs)

Document modeling
- Observed: words.
- Hidden: topics.

Social Network Modeling
- Observed: social interactions.
- Hidden: communities, relationships.

Recommendation Systems
- Observed: recommendations (e.g., reviews).
- Hidden: User and business attributes

Unsupervised Learning: Learn LVM without labeled examples.
LVM for Feature Engineering

- Learn good features/representations for classification tasks, e.g., computer vision and NLP.

Sparse Coding/Dictionary Learning

- **Sparse** representations, low dimensional hidden structures.
- A few **dictionary** elements make complicated shapes.
Challenges in Learning LVMs

Computational Challenges

- Practice: Local search approaches such as gradient descent, EM, Variational Bayes have no consistency guarantees.
- Can get stuck in bad local optima. Poor convergence rates.
- Hard to parallelize

Alternatives? Guaranteed and efficient learning?
Classical Spectral Methods: Matrix PCA

For centered samples \( \{x_i\} \), find projection \( P \) with \( \text{Rank}(P) = k \) s.t.

\[
\min_{P} \frac{1}{n} \sum_{i \in [n]} \|x_i - Px_i\|^2.
\]

Result: Eigen-decomposition of \( \text{Cov}(X) \).

Beyond PCA: Spectral Methods on Tensors?
Moment Matrices and Tensors

Multivariate Moments

\[ M_1 := \mathbb{E}[x], \quad M_2 := \mathbb{E}[x \otimes x], \quad M_3 := \mathbb{E}[x \otimes x \otimes x]. \]

Matrix

- \( \mathbb{E}[x \otimes x] \in \mathbb{R}^{d \times d} \) is a second order tensor.
- \( \mathbb{E}[x \otimes x]_{i_1,i_2} = \mathbb{E}[x_{i_1} x_{i_2}] \).
- For matrices: \( \mathbb{E}[x \otimes x] = \mathbb{E}[xx^\top] \).

Tensor

- \( \mathbb{E}[x \otimes x \otimes x] \in \mathbb{R}^{d \times d \times d} \) is a third order tensor.
- \( \mathbb{E}[x \otimes x \otimes x]_{i_1,i_2,i_3} = \mathbb{E}[x_{i_1} x_{i_2} x_{i_3}] \).
Spectral Decomposition of Tensors

\[ M_2 = \sum_i \lambda_i u_i \otimes v_i \]

Matrix \( M_2 \)

\( \lambda_1 u_1 \otimes v_1 \) + \( \lambda_2 u_2 \otimes v_2 \)
Spectral Decomposition of Tensors

Matrix $M_2$

$$M_2 = \sum_i \lambda_i u_i \otimes v_i$$

$$= \lambda_1 u_1 \otimes v_1 + \lambda_2 u_2 \otimes v_2$$

$u \otimes v \otimes w$ is a rank-1 tensor since its $(i_1, i_2, i_3)^{th}$ entry is $u_{i_1} v_{i_2} w_{i_3}$.

Efficient methods for tensor decomposition
Topic Modeling

$k$ topics (distributions over vocab words).

Each document $\leftrightarrow$ mixture of topics.

Words in document $\sim_{iid}$ mixture dist.

E.g.,

<table>
<thead>
<tr>
<th></th>
<th>aardvark</th>
<th>athlete</th>
<th>zygote</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

$0.6 \cdot \text{sports} + 0.3 \cdot \text{science} + 0.1 \cdot \text{politics} + 0.0 \cdot \text{business}$

$\Pr_{\theta}[\text{“play” | sports}] = 0.0002$
$\Pr_{\theta}[\text{“game” | sports}] = 0.0003$
$\Pr_{\theta}[\text{“season” | sports}] = 0.0001$
Tractable Learning for LVMs

GMM

Multiview and Topic Models

HMM

ICA

Multiview and Topic Models

\[ h_1, h_2, h_3 \]

\[ x_1, x_2, x_3 \]

\[ h_1, h_2, \ldots, h_k \]

\[ x_1, x_2, \ldots, x_d \]

\[ h \in [k], \]

\[ x_1 \in \mathbb{R}^{d_1}, x_2 \in \mathbb{R}^{d_2}, \ldots, x_\ell \in \mathbb{R}^{d_\ell}. \]

\[ k = \text{# components}, \ \ell = \text{# views (e.g., audio, video, text)}. \]

View 1: \( x_1 \in \mathbb{R}^{d_1} \)

View 2: \( x_2 \in \mathbb{R}^{d_2} \)

View 3: \( x_3 \in \mathbb{R}^{d_3} \)
Overall Framework

Unlabeled Data → Probabilistic admixture models → Tensor Method =  + ... → Inference
Conclusion: Tensor Methods for Learning

Tensor Decomposition
- Efficient sample and computational complexities
- Better performance compared to EM, Variational Bayes etc.

In practice
- Scalable and embarrassingly parallel: handle large datasets.
- Efficient performance: perplexity or ground truth validation.

Related Topics
- Tensor Methods for Discriminative Learning: Learning neural networks, mixtures of classifiers, etc.
- Overcomplete Tensor Decomposition: Neural networks, sparse coding and ICA models tend to be overcomplete (more neurons than input dimensions).
My Research Group and Resources

Furong Huang  Majid Janzamin  Hanie Sedghi

Niranjan UN  Forough Arabshahi

ML summer school lectures available at
http://newport.eecs.uci.edu/anandkumar/MLSS.html