Low Rank Approximations of Tensors

Presenter: Richard Peng

Mar 2, 2017

**DISCLAIMER:** These notes are not necessarily an accurate representation of what I said during the class. They are mostly what I intend to say, and have not been carefully edited.

Today we will talk about generalizations of low rank approximations to higher dimensions, specifically $n^r$ sized data tensors. What we give is in fact another algorithm for computing singular vectors though.

- **Tensors:** $A \in \mathbb{R}^{n^r}$:
  
  \[ A(x^{(1)}, \ldots x^{(r)}) = \sum_{i_1, i_2, \ldots, i_r} A_{i_1 i_2 \ldots i_r} \cdot x_{i_1}^{(1)} \cdot x_{i_2}^{(2)} \cdot \ldots \cdot x_{i_r}^{(r)}. \]

- **Interpretation:** transition probabilities involving $r - 1$ previous states.

- **Frobenius norm:** sum over square of all $n^2$ entries.

- **Spectral norm:**
  \[
  \max_{\|z_2^{(1)}\| = 1, \|z_2^{(2)}\| = 1, \ldots, \|z^{(r)}\| = 1} A(z^{(1)}, \ldots z^{(r)}).
  \]

- **Rank-1 tensor:**
  
  \[ x^{(1)} \otimes x^{(r)}, \]

  entry $i_1 i_2 \ldots i_r$ given by

  \[ x_{i_1}^{(1)} \cdot x_{i_2}^{(2)} \cdot \ldots \cdot x_{i_r}^{(r)}. \]

- **Aside:** $A(x^{(1)}, \ldots x^{(r)})$ can be viewed as the ‘dot product’ between $A$ and $x^{(1)} \otimes x^{(2)} \otimes \ldots \otimes x^{(r)}$.

- **Main goal:** for some error $\epsilon$, find rank-1 tensors $B_1 \ldots B_k$ such that
  \[ \|A - B_1 - \ldots - B_k\|_2 \leq \epsilon \|A\|_F. \]

- **Algorithm:** while $\|A - B_1 - \ldots - B_k\|_2 > \epsilon \|A\|_F$, create $B_{k+1}$ from the vectors $x^{(1)} \otimes x^{(2)} \otimes \ldots \otimes x^{(r)}$ that maximize this value. Can show: terminates in $k = O(\epsilon^{-2})$ steps.

- **Rest of this lecture:** find unit vectors $z^{(1)} \ldots z^{(r)}$ that maximize $A(z^{(1)} \ldots z^{(r)})$. 
• Observation: if \( z^{(1)} \ldots z^{(r-1)} \) is fixed, then \( z^{(r)} \) should be set to the normalized version of the vector \( A(z^{(1)}, \ldots, z^{(r-1)}, \cdot) \).

• Details of the algorithm is in the notes, but the 2-D version is:
  
  – Pick about \( O(r^3 \epsilon^{-2}) \) random ‘rows’ from \( A \), with probability proportional to the sum of squares, 
    \[ \sum_i A_{i_1, i_2, \ldots, i_{r-1}, i}^2. \]
  
  – Enumerate values of \( z^{(1)}_{i_1} \ldots z^{(r-1)}_{i_{r-1}} \) of those entries from a set of coordinates with granularity \( n^{-1/2} \).
  
  – Compute the vector in the last dimension (\( r \)) using this set of sampled values: for each \( 1 \leq i \leq n \), set
    \[ y_i \leftarrow \sum_{i_1, i_2, \ldots, i_{r-1} \in I} A(i_1, i_2, \ldots, i_{r-1}, i) \hat{z}^{(1)}_{i_1} \hat{z}^{(2)}_{i_2} \ldots \hat{z}^{(r-1)}_{i_{r-1}}. \]
  
  – Build an \( r-1 \) dimensional tensor using this vector \( y \), recurse on it to find the best \( r-1 \)-dimensional eigenvector. Return maximum among all the enumerated values.

• Proving the correctness of this has several key steps:
  
  – Showing that rounding all entries to nearest multiple copies of \( n^{-1/2} \) changes the value by at most \( \epsilon \| A \|_F \).
  
  – Showing that any such \( z^{(1)} \ldots z^{(r-1)} \), computing values of \( A(z^{(1)} \ldots z^{(r-1)}, \cdot) \) via random sampling gives an unbiased estimator whose variance is bounded by \( \| A \|_F^2 \).
  
  – Combining these to show that among the things enumerated, there exists a vector \( y \) such that
    \[ \| A(z^{(1)} \ldots z^{(r)}) - A(z^{(1)} \ldots z^{(r-1)}, y) \|_F \leq \frac{\epsilon}{10r} \| A \|_F \]
    
    with probability at least \( 1 - 1/10r \).
  
  – Applying this statement inductively to the next level shows that the error compounds to at most \( \epsilon \) with constant probability.