Fast Tucker Rank Reduction for Non-Negative Tensors Using Mean-Field Approximation

Kazu Ghalamkari 1,2, Mahito Sugiyama 1,2

1National Institute of Informatics @ Tokyo, Japan 2The Graduate University for Advanced Studies, SOKENDAI

The 34th Annual Conference on Neural Information Processing Systems (NeurIPS 2021), December 6–14, 2021

Summary

- Low-rank tensors reduce memory requirements.
- Many non-negative low-rank approximation methods are based on a gradient method.
  ⇒ Initial values, stopping criterion, learning rate...

We developed fast low-rank approximation without gradient method, called Legendre Rank Reduction (LTR).

Experiments

- LTR is Faster.
- LTR has Competitive error.

Theory

- Tensor as distribution
  \[ p_\theta(i,j,k) = \exp \left( \sum_{I=1}^{K} \sum_{J=1}^{K} \sum_{K=1}^{K} \theta_{i'j'k'} \right) \]
  \[ \eta_{ijk} = \sum_{i'=1}^{K} \sum_{j'=1}^{K} \sum_{k'=1}^{K} \mathcal{P}_{i'j'k'} \]
  - \( \theta \)-representation
  - \( \eta \)-representation

- Describe rank-1 condition using \( (\theta, \eta) \)
  - Rank-1 condition (\( \eta \)-representation)
    \[ \text{rank}(\mathcal{P}) = 1 \iff \eta_{ijk} = \eta_{i11j11k11} \]
  - Rank-1 condition (\( \theta \)-representation)
    \[ \text{rank}(\mathcal{P}) = 1 \iff \text{its all many-body} \theta \text{ parameters are} 0 \]

- Best rank-1 tensor formula for minimizing KL divergence
  For any positive tensor \( \mathcal{P} \), its best rank-1 approximation is
  \[ \mathcal{P}_{i'j'k'} = \left( \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{k=1}^{K} \mathcal{P}_{ijk} \right) \]
  \[ = \left( \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{k=1}^{K} \mathcal{P}_{ijk} \right) \]

Legendre Rank Reduction

- Let us reduce a \((8,8,3)\)-tensors rank to \((5,7,3)\).
  - Low-rank approximation is a projection onto Bingo space.
  - \( \mathcal{P} \) is the best tensor in the specified bingo space.

Step 1: Choose a bingo location, randomly.
- \( \theta \) can be any value
- \( \theta \) is zero

Step 2: Replace the bingo part with the rank-1 tensor.

Step 3: Do Step 1 and Step 2 in other axes.

The shaded areas do not change their values in the projection.

Rank-1 Tensor

Rank-1 Tensor

Bingo

Bingo

Bingo

Two bingos

Two bingos

The mode-k expansion \( \theta^{(k)} \) of the natural parameter has \( b_k \) bingos
\[ \Rightarrow \text{rank}(\mathcal{P}) \leq (l - b_1) + b_2 + K - b_3 \]