Tensor (an object of multilinear algebra)

Tensor A is a mapping \( A : I \to \mathbb{R} \), where
- \( I = i_1 \times \ldots \times i_k \),
- \( k \) is a natural number called the order of tensor A,
- \( i_j, j = 1, \ldots, k \) are index sets (typically, sets of integers),
- cardinalities of \( i_j, j = 1, \ldots, k \) are called dimensions of tensor A.

Example

Example of a tensor of order \( k = 4 \) and dimensions \( n_1 = n_2 = n_3 = n_4 = 2 \):

\[
A = \begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1
\end{pmatrix}
\]

Tensor rank

Tensor A has rank one if it can be written as an outer product of vectors:

\[
A = a_1 \otimes \ldots \otimes a_k,
\]

with the outer product \( \otimes \) being defined as

\[
A_{i_1 \ldots i_k} = a_{i_1} \ldots a_{i_k},
\]

for all \( (i_1, \ldots, i_k) \in I_1 \times \ldots \times I_k \).

Each tensor can be decomposed as a linear combination of rank-one tensors:

\[
A = \sum_{r=1}^{r_k} b_r \cdot a_{r,1} \otimes \ldots \otimes a_{r,k},
\]

The rank of a tensor A, denoted \( \text{rank}(A) \), is the minimal \( r \) over all such decompositions. The decomposition of a tensor A to tensors of rank one that sum up to A is called CP tensor decomposition.

Example

The tensor A from the above example can be written as:

\[
A = (1, 1) \otimes (1, 1) \otimes (1, 1) - (1, 0) \otimes (1, 0) \otimes (1, 0) .
\]

This implies that its rank is at most 2.

Conditional probability tables (CPTs) as tensors

- Let \( X_i, i = 1, \ldots, k \) be discrete random variables taking states from \( I_j, j = 1, \ldots, k \) and
- let \( Y \) be a discrete random variable with an observed state \( y \).

Then each CPT \( P(Y = y | X_1, \ldots, X_k) \) can be viewed as a tensor.

Noisy-or: an example of a CPT of a low rank

- Let \( X_i, i = 1, \ldots, k \) and \( X'_i, i = 1, \ldots, k \) be binary variables taking states from \( \{0, 1\} \).
- \( Y \) takes value 1 if at least one of its parents take value 1:

\[
P(Y = 1 | X'_1, \ldots, X'_k = x'_k) = \begin{cases} 1 & \text{if } x'_1 + \ldots + x'_k > 0 \\ 0 & \text{otherwise.} \end{cases}
\]

- For \( i = 1, \ldots, k \):

\[
P(X'_i = 1 | X_i = x) = \begin{cases} 0 & \text{if } x = 0 \\ \pi_i & \text{otherwise.} \end{cases}
\]

Example

The CPT of \( P(Y = 1 | X'_1 = x'_1, \ldots, X'_k = x'_k) \) corresponds to the tensor A from the above example.

Probabilistic inference

Compute \( P(X|e) \) where \( e \) is the collected evidence.

Example

Compute \( P(X|Y = y) \) for \( j = 1, \ldots, 4 \).

Low rank means efficient inference

A measure of the complexity of probabilistic inference is the total table size (tts). It is the total size of the tables used in a probabilistic inference method applied to a Bayesian network.

Example

Bayesian network with one noisy-or:

\[
P(Y = y | X_1, \ldots, X_k) \]

\[
\sum_{i=1}^{r_k} \psi_i(X_i, Y) \otimes \psi_i(X_i, Y)
\]

Real data and algorithms used

- 15 Bayesian networks from a Bayesian network repository: alarm.net, hailfinder.net, barley2.net, pathfinder.net, munin1.net, munin2.net, munin3.net, munin4.net, mildew2.net, hepar2.net, andes.net, win95pts.net, water.net, link.net, and insurance.net.
- In the experiments we considered all CPTs for variables having at least three parents. It was 492 CPTs in total.
- For the CP tensor decomposition we used the Fast Damped Gauss-Newton (Levenberg-Marquard) algorithm implemented in the Matlab package TENSORBOX.

How common are low rank CPTs in real applications?

Percentage of CPTs that can be approximated with a maximum error smaller than 0.001 as a function of the rank of the approximation.

Conclusions

- Most of the conditional probability tables can be very well approximated by tables that have lower rank than one would expect from a general table of the same dimensions.
- The low rank approximation should be exploited (1) in model elicitation by using a compact parametrization of the internal structure of a CPT but also (2) during probabilistic inference.
- We conjecture that some low rank approximations of CPTs may actually correspond better to what a domain expert intended to model in the constructed CPT.