Learning Predictors from Multidimensional Data with Tensor Factorizations

Soo Min Kwon*, Anand D. Sarwate*

*Rutgers, The State University of New Jersey

Motivations & Objectives

Motivations:

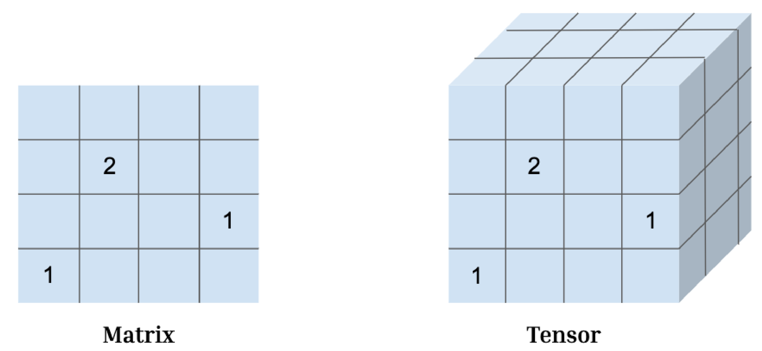
- Traditional machine learning algorithms are formulated for vectors as inputs
- Vectorizing multidimensional data to fit these algorithms can be computationally expensive
- Vectorizing data destroys inherent spatial and temporal structure of the tensor

Objectives:

- Formulate structured machine learning algorithms that exploits structure of tensor data
- Impose tensor factorizations (e.g. CANDECOMP/PARAFAC decomposition) on predictors to reduce the parameters to be estimated

Preliminaries

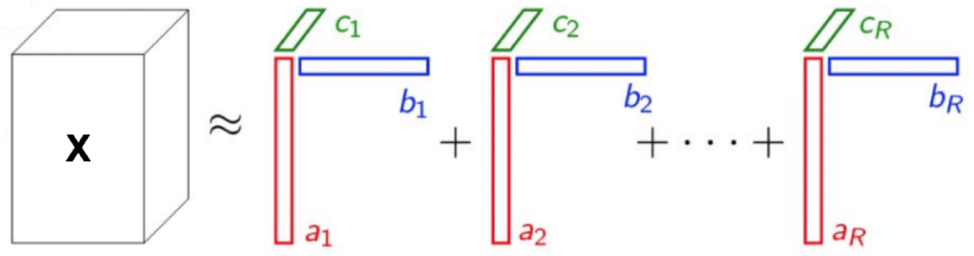
Tensors:



Tensors are multidimensional arrays:

- Formally denoted as $\mathbf{X} \in \mathbb{R}^{D_1 \times D_2 \times \cdots \times D_N}$
- Generalizations of vectors and matrices
- Useful for modelling data with many variables

CANDECOMP/PARAFAC (CP) Decomposition:



CP decomposition expresses a tensor as the sum of component rank–one tensors [1]. For a three-dimensional tensor, this is formulated as:

$$\mathbf{X} \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r. \tag{1}$$

Methodology

Consider the ERM problem for two types of classifiers:

Support Vector Machines (SVM):

$$\underset{w}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} \max[0, 1 - y_i(w^\top x_i)] + \frac{\lambda}{2}||w||^2. \tag{2}$$

Logistic Regression (LOGIT):

$$\underset{w}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i(w^\top x_i))) + \lambda||w||^2. \tag{3}$$

Problem Formulation:

Consider $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$, where $\mathbf{X}_i \in \mathbb{R}^{D_1 \times \ldots \times D_N}$ denotes a tensor data sample with $y_i \in \{-1, 1\}$. By imposing (1) onto predictors $w^\top$ of (2) and (3):

CP Structured SVM:

$$\underset{W_1,\ldots,W_N}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} \max[0, 1 - y_i(\langle \sum_{r=1}^{R} W_1^{(r)} \circ W_2^{(r)} \circ \ldots \circ W_N^{(r)}, \mathbf{X}_i \rangle)] \tag{4}$$

CP Structured LOGIT:

$$\underset{W_1,\ldots,W_N}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i(\langle \sum_{r=1}^{R} W_1^{(r)} \circ W_2^{(r)} \circ \ldots \circ W_N^{(r)}, \mathbf{X}_i \rangle))) \tag{5}$$

Remarks:

- Our models significantly reduces the number of parameters to be estimated
- Can solve for factor matrices $W_1, W_2, \ldots, W_N$ using an alternating minimization method as proposed by Zhou et al. [2]

Conclusion & Future Work

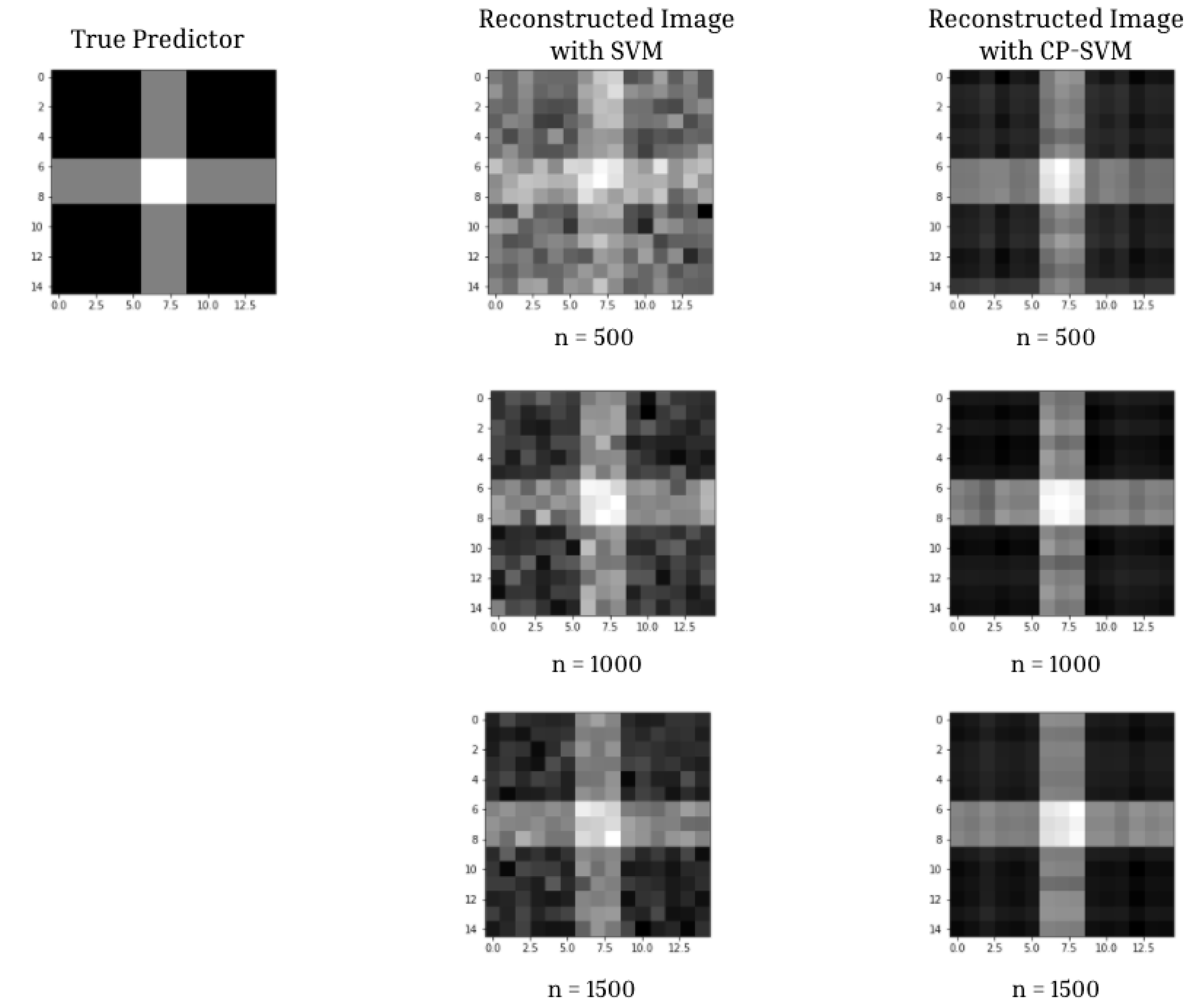
- Imposing structure on predictors allows us to solve for less parameters more efficiently
- Structured algorithms works efficiently when true tensor predictor exhibits low–rank structure
- Performance of traditional algorithm converges to structured algorithm as sample size increases
- This is a **proof of concept** – future work involves finding real datasets to test these methods

Results

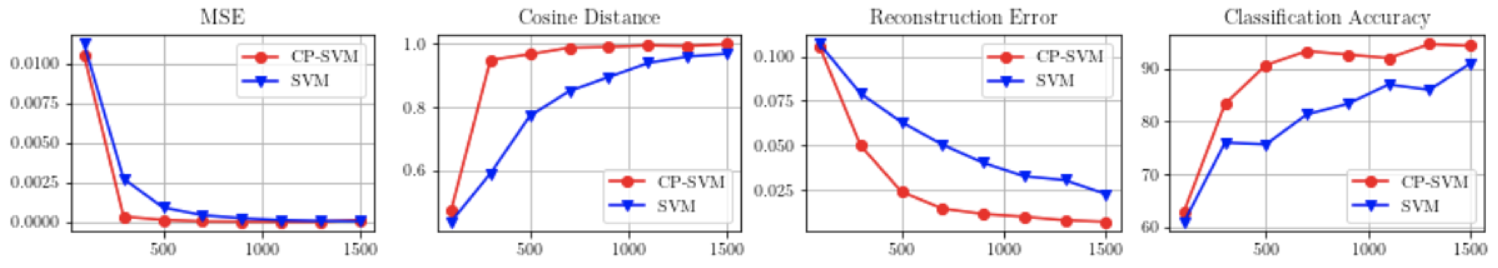
Goal: Fix predictor to be a (15×15) image and generate i.i.d Gaussian data (X_i, y_i)

- See if our models can estimate true predictor given (X_i, y_i)
- Evaluate performance of models using metrics MSE, cosine distance, reconstruction error, and classification accuracy

Results (for different sample sizes):



Performance Metrics (with increasing sample sizes):



References

[1] T. Kolda and B. Bader, "Tensor decompositions and applications", SIAM Review,vol. 51, pp. 455â500, 08 2009.

[2] H. Zhou, L. Li, and H. Zhu, "Tensor regression with applications in neuroimaging dataanalysis", Journal of the American Statistical Association, vol. 108, pp. 540â552, 062013.