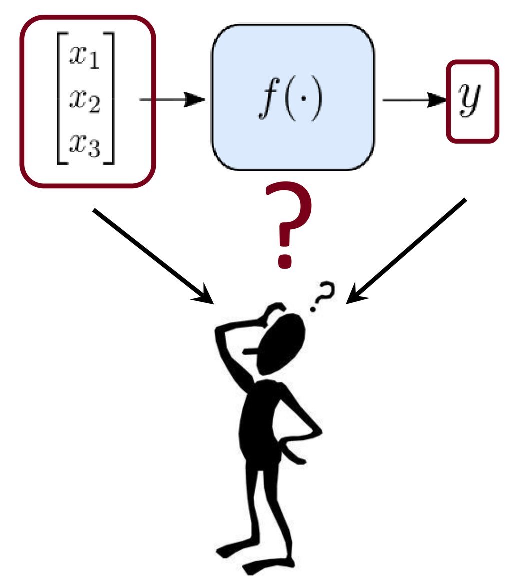
NONLINEAR SYSTEM IDENTIFICATION VIA TENSOR COMPLETION

Nikos Kargas and Nicholas D. Sidiropoulos

University of Minnesota
Driven to Discover
umn.edu

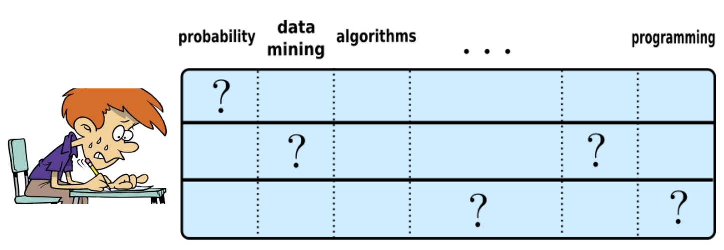
The Supervised Learning Problem

- General nonlinear function identification
  - `Supervised' - from input-output data
  - Function approximation problem
  - Identifiability? Performance? Complexity?

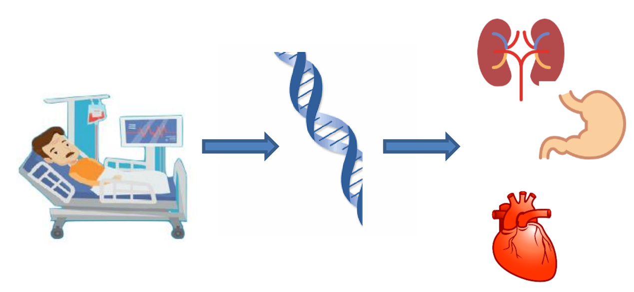


Categorical (classification)
Real-valued (regression)

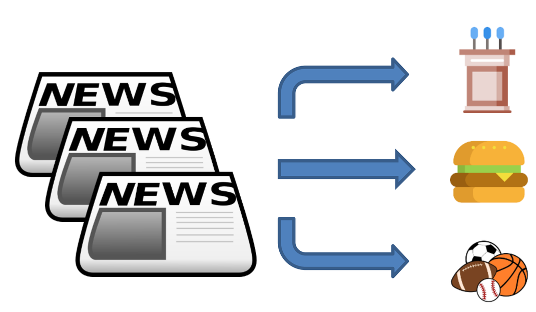
- Applications
  - Machine learning
  - Dynamical system identification and control
  - Communications



Course grade prediction

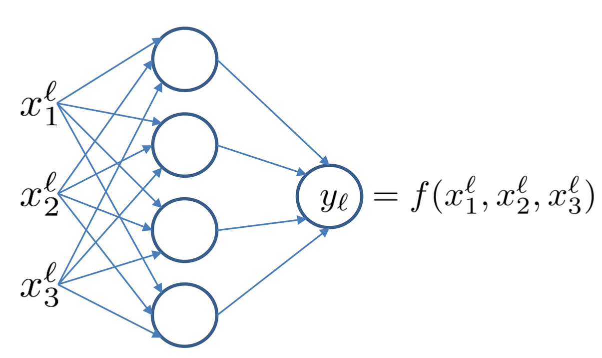


Drug response prediction



Text classification

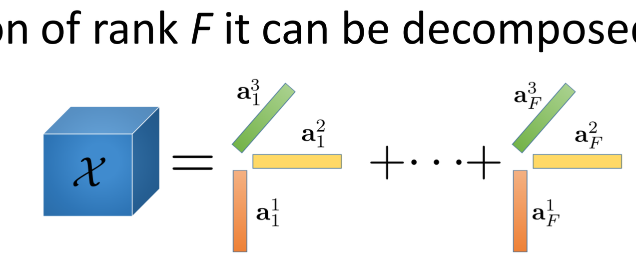
- Neural Networks
  - Most popular method for learning to mimic nonlinear functions
  - Work very well in practice
  - Don't understand why they work so well
  - Choosing architecture is art
  - Hard to interpret



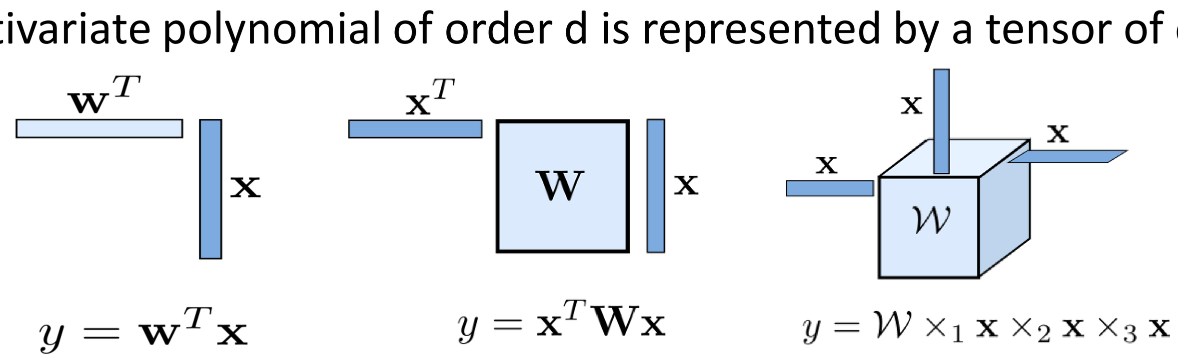
Background

- Canonical Polyadic Decomposition (CPD)
  - An N-way tensor admits a decomposition of rank *F* it can be decomposed as a sum of *F* rank-1 tensors

$$\mathcal{X} = \sum_{f=1}^{F} \mathbf{a}_f^1 \circ \mathbf{a}_f^2 \circ \cdots \circ \mathbf{a}_f^N$$



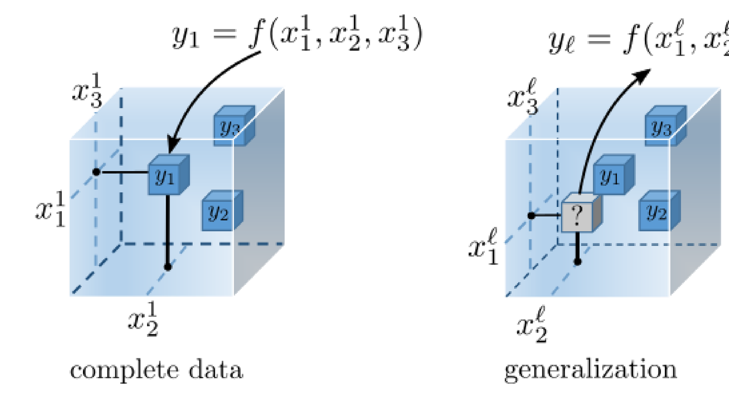
  - Tensor rank is smallest *F* for which such decomposition exists → *Canonical*
  - Element-wise: $\mathcal{X}(i_1, \ldots, i_N) = \sum_{f=1}^{F} \prod_{n=1}^{N} \mathbf{a}_f^n(i_n)$
  - Matrix unfolding: $\mathcal{X}^{(n)} = (\mathbf{A}_N \odot \cdots \odot \mathbf{A}_{n+1} \odot \mathbf{A}_{n-1} \cdots \odot \cdots \mathbf{A}_1)\mathbf{A}_n^T$
  - Vector $\mathrm{vec}(\mathcal{X}) = (\mathbf{A}_N \odot \cdots \odot \mathbf{A}_1)\mathbf{1}$
- Prior Work
  - Tensor modeling of low-order multivariate polynomial systems
  - A multivariate polynomial of order d is represented by a tensor of order d



- Drawbacks
  - Require prior knowledge of polynomial order
  - Assuming polynomial of a given degree can be restrictive
  - Simplest rank=1 model ➡ number of parameters grows linearly with d
  - Cannot model high-degree polynomial functions

Canonical System Identification (CSID)

- We propose:
  - Single high-order tensor for learning a general nonlinear system



- Claims:
  - CPD can model *any* nonlinearity (even of ∞ order) for high-enough rank. Even for low ranks, it can model highly nonlinear operators
  - Provably correct nonlinear system identification from limited samples, when the tensor is low rank
  - Even when not low rank ➡ identification of the principal components!

- Rank of Generic Nonlinear Systems
  - Separable function: $y = f(x_1, \ldots, x_N) = \prod_{n=1}^{N} f_n(x_n)$
    - Rank: 1, e.g., $f(x_1, \ldots, x_N) = \prod_{n=1}^{N} \mathrm{sign}(x_n)$
  - Sum of separable functions: $y = f(x_1, \ldots, x_N) = \sum_{n=1}^{N} f_n(x_n)$
    - Maximal rank: N, e.g., $f(x_1, \ldots, x_N) = \sum_{n=1}^{N} \mathrm{sign}(x_n)$
  - Sum of pairwise functions $y = f(x_1, \ldots, x_N) = \sum_{i=1}^{N} \sum_{j>i} f_{ij}(x_i, x_j)$
    - Maximal rank: $\frac{IN^2}{2} \ll I^{N-1}$

Problem Formulation

- Smooth Tensor Completion

$$\min_{\mathcal{X}, \{\mathbf{A}_n\}_{n=1}^N} \frac{1}{M} \left\| \sqrt{\mathcal{W}} \circledast (\mathcal{Y} - \mathcal{X}) \right\|_F^2 + \sum_{n=1}^{N} \rho \|\mathbf{A}_n\|_F^2 + \sum_{n=1}^{N} \mu_n \|\mathbf{T}_n \mathbf{A}_n\|_F^2$$

$$\text{subject to} \quad \mathcal{X} = \sum_{f=1}^{F} \mathbf{A}_1(:, f) \odot \cdots \odot \mathbf{A}_N(:, f),$$

$$\mathbf{T}_n = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \quad \text{or} \quad \mathbf{T}_n = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}$$

- Alternating minimization
  - Exploit sparsity
  - Cyclically update variables
  - Lightweight row-wise updates

Experimental Results

- Baselines: Ridge Regression (RR), Support Vector Regression (SVR), Decision Tree (DT), Neural network: multilayer perceptron (MLP)

Dataset	RR	SVR (RBF)	SVR (polynomial)	DT	MLP (5 Layer)	CSID
Energy Eff. (1)	2.91±0.17	2.68±0.17	4.09±0.49	0.56±0.03	**0.48±0.06** [50]	**0.39±0.05**
Energy Eff. (2)	3.09±0.19	3.03±0.21	4.14±0.44	1.86±0.19	**0.97±0.14** [50]	**0.57±0.09**
C. Comp. Strength	10.47±0.42	9.72±0.38	11.30±0.36	6.57±0.82	**4.92±0.63** [50]	**4.67±0.50**
SkillCraft Master Table	1.68±1.61	**0.99±0.03**	1.22±0.05	1.03±0.04	1.00±0.03 [10]	**0.91±0.02**
Abalone	2.25±0.10	**2.19±0.08**	3.90±3.43	2.35±0.08	**2.09±0.09** [10]	2.23±0.09
Wine Quality	0.76±0.02	**0.69±0.02**	1.01±0.39	0.75±0.03	0.72±0.02 [10]	**0.70±0.02**
Parkinsons Tel. (1)	7.51±0.11	6.66±0.14	7.89±0.88	**2.40±0.26**	3.60±0.18 [100]	**1.33±0.10**
Parkinsons Tel. (2)	9.75±0.15	9.14±0.17	10.04±0.43	**2.60±0.38**	5.01±0.19 [100]	**1.79±0.17**
C. Cycle Power Plant	5.51±0.09	4.13±0.09	8.00±0.19	**3.98±0.13**	4.06±0.11 [50]	**3.76±0.15**
Bike Sharing (1)	36.45±0.46	32.67±0.81	34.93±0.97	18.89±0.36	**14.81±0.44** [100]	**15.17±0.44**
Bike Sharing (2)	122.65±2.87	113.18±1.73	117.25±2.01	42.06±2.06	**38.69±1.24** [100]	**36.93±1.19**
Phys. Prop.	5.19±0.03	4.91±1.26	6.49±1.15	4.40±0.04	**4.20±0.05** [100]	**4.21±0.04**

- Grade Prediction

Dataset	GPA	BMF	CSID
CSCI-1	0.52±0.02	**0.48±0.03**	**0.48±0.03**
CSCI-2	0.56±0.02	**0.55±0.02**	**0.55±0.03**
CSCI-3	**0.48±0.04**	**0.48±0.04**	**0.48±0.05**
CSCI-4	0.53±0.03	0.52±0.04	**0.51±0.03**
CSCI-5	0.43±0.02	0.43±0.02	**0.42±0.02**
CSCI-6	0.63±0.03	0.58±0.03	**0.57±0.03**
CSCI-7	0.57±0.02	0.58±0.01	**0.56±0.02**
CSCI-8	0.52±0.02	0.49±0.03	**0.47±0.02**
CSCI-9	0.61±0.03	0.60±0.05	**0.57±0.03**
CSCI-10	0.58±0.04	**0.56±0.04**	**0.56±0.04**

Dataset	GPA	BMF	CSID
CSCI-11	0.68±0.06	**0.66±0.04**	0.67±0.03
CSCI-12	0.58±0.04	0.51±0.04	**0.48±0.01**
CSCI-13	0.67±0.03	0.55±0.05	**0.54±0.03**
CSCI-14	0.70±0.06	**0.62±0.03**	0.65±0.07
CSCI-15	0.56±0.03	**0.56±0.06**	0.57±0.03
CSCI-16	0.52±0.03	0.51±0.03	**0.50±0.02**
CSCI-17	0.60±0.02	**0.58±0.05**	0.59±0.05
CSCI-18	0.57±0.03	0.56±0.05	**0.55±0.04**
CSCI-19	0.68±0.04	0.70±0.04	**0.61±0.04**
CSCI-20	0.61±0.06	**0.58±0.02**	0.63±0.04

University of Minnesota
Driven to Discover
umn.edu