Low-Rank Tensor Regression

Jean Kossaifi

@JeanKossaifi

jean.kossaifi@gmail.com

Tensor Regression

- Classical regression loss: $\frac{1}{2}\sum_{i=1}^{N} \| y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$

$$
\boxed{\begin{aligned} < \hat{\mathscr{X}}, \hat{\mathscr{Y}} > &= \sum_{i_0=0}^{I_0} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=0}^{I_N} x_{i_0 i_1 \cdots i_N} y_{i_0 i_1 \cdots i_N} \\ &= vec(\hat{\mathscr{X}})^\top vec(\hat{\mathscr{Y}}) \end{aligned}}
$$

@JeanKossaifi

Tensor Regression

- Classical regression loss: $\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathcal{X}}_i, \hat{\mathcal{W}} > \|_F^2$

- Note that this is equivalent to linear regression as:

$$< \hat{\mathcal{X}}_i, \hat{\mathcal{W}} > = vec(\hat{\mathcal{X}}_i)^\top vec(\hat{\mathcal{W}})$$

Tensor Regression

- Classical regression loss: $\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$

- Note that this is equivalent to linear regression as:

$$< \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > = vec(\hat{\mathscr{X}}_i)^\top vec(\hat{\mathscr{W}})$$

- Difference comes from imposing low-rank structure on $\hat{\mathscr{W}}$

Tensor Regression

- Classical regression loss: $\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$

- Note that this is equivalent to linear regression as:

$$< \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > = vec(\hat{\mathscr{X}}_i)^\top vec(\hat{\mathscr{W}})$$

- Difference comes from imposing low-rank structure on $\hat{\mathscr{W}}$

$$\hat{\mathscr{W}} = [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,]$$

✔ **Kruskal form (CP) but can be any tensor decomposition**

@JeanKossaifi

Tensor Regression

- Classical regression loss: $\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$

- Note that this is equivalent to linear regression as:

$$< \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > = vec(\hat{\mathscr{X}}_i)^\top vec(\hat{\mathscr{W}})$$

- Difference comes from imposing low-rank structure on $\hat{\mathscr{W}}$

$$\hat{\mathscr{W}} = [\![\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}]\!]$$

- Structure in the regression weights / less parameters

@JeanKossaifi

Tensor Regression

$$\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$$

@JeanKossaifi

Tensor Regression

$$\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \underbrace{[\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,]}_{\textbf{Kruskal tensor}} > \|_F^2$$

@JeanKossaifi

Tensor Regression

$$\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \underbrace{[\![\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}]\!]}_{\textbf{Kruskal tensor}} > \|_F^2$$

- Add regularisation on the weights

$$\frac{\lambda}{2}\|\hat{\mathscr{W}}\|_F^2$$

Tensor Regression

$$\frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \hat{\mathscr{W}} > \|_F^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \|y_i - < \hat{\mathscr{X}}_i, \underbrace{[\![\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}]\!]}_{\textbf{Kruskal tensor}} > \|_F^2$$

- Add regularisation on the weights

$$\cancel{\frac{\lambda}{2}\|\hat{\mathscr{W}}\|_F^2} \qquad \frac{\lambda}{2}\sum_{k=0}^{N}\|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

  - Separable regularisation term
  - Closed-form solution

@JeanKossaifi

Tensor Regression

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - < \hat{\mathscr{X}}_i, [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,] > \|_F^2 + \frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

Tensor Regression

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - < \hat{\mathscr{X}}_i, [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,] > \|_F^2 + \frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - Trace\left(\mathbf{U}^{(n)} \underbrace{\left(\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \cdots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \cdots \odot \mathbf{U}^{(N)}\right)^{\top} (\hat{\mathscr{X}}_i)_{[n]}^{T}}_{\mathbf{\Phi}^{\top}}\right) \|_F^2$$

$$+\frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

$$\boxed{\left(\textbf{Trace}(\mathbf{A}\mathbf{B}^{\top}) = vec(\mathbf{A})^{\top} vec(\mathbf{B})\right)}$$

Tensor Regression

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - < \hat{\mathscr{X}}_i, [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,] > \|_F^2 + \frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - Trace\left(\mathbf{U}^{(n)} \underbrace{\left(\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \cdots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \cdots \odot \mathbf{U}^{(N)}\right)^{\top} (\hat{\mathscr{X}}_i)_{[n]}^T}_{\boldsymbol{\Phi}^{\top}}\right) \|_F^2$$

$$+\frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

- Build $\boldsymbol{\Phi}$ so that row k contains

$$(\hat{\mathscr{X}}_i)_{[n]}\left(\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \quad \cdots \quad \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \quad \cdots \quad \odot \mathbf{U}^{(N)}\right)$$

@JeanKossaifi

Tensor Regression

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - < \hat{\mathscr{X}}_i, [\,|\,\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\,|\,] > \|_F^2 + \frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

$$\frac{1}{2}\sum_{i=0}^{N} \|y_i - Trace\left(\mathbf{U}^{(n)} \underbrace{\left(\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \cdots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \cdots \odot \mathbf{U}^{(N)}\right)^{\top} (\hat{\mathscr{X}}_i)^T_{[n]}}_{\mathbf{\Phi}^{\top}}\right) \|_F^2$$

$$+\frac{\lambda}{2}\sum_{k=0}^{N} \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

- Build $\mathbf{\Phi}$ so that row k contains

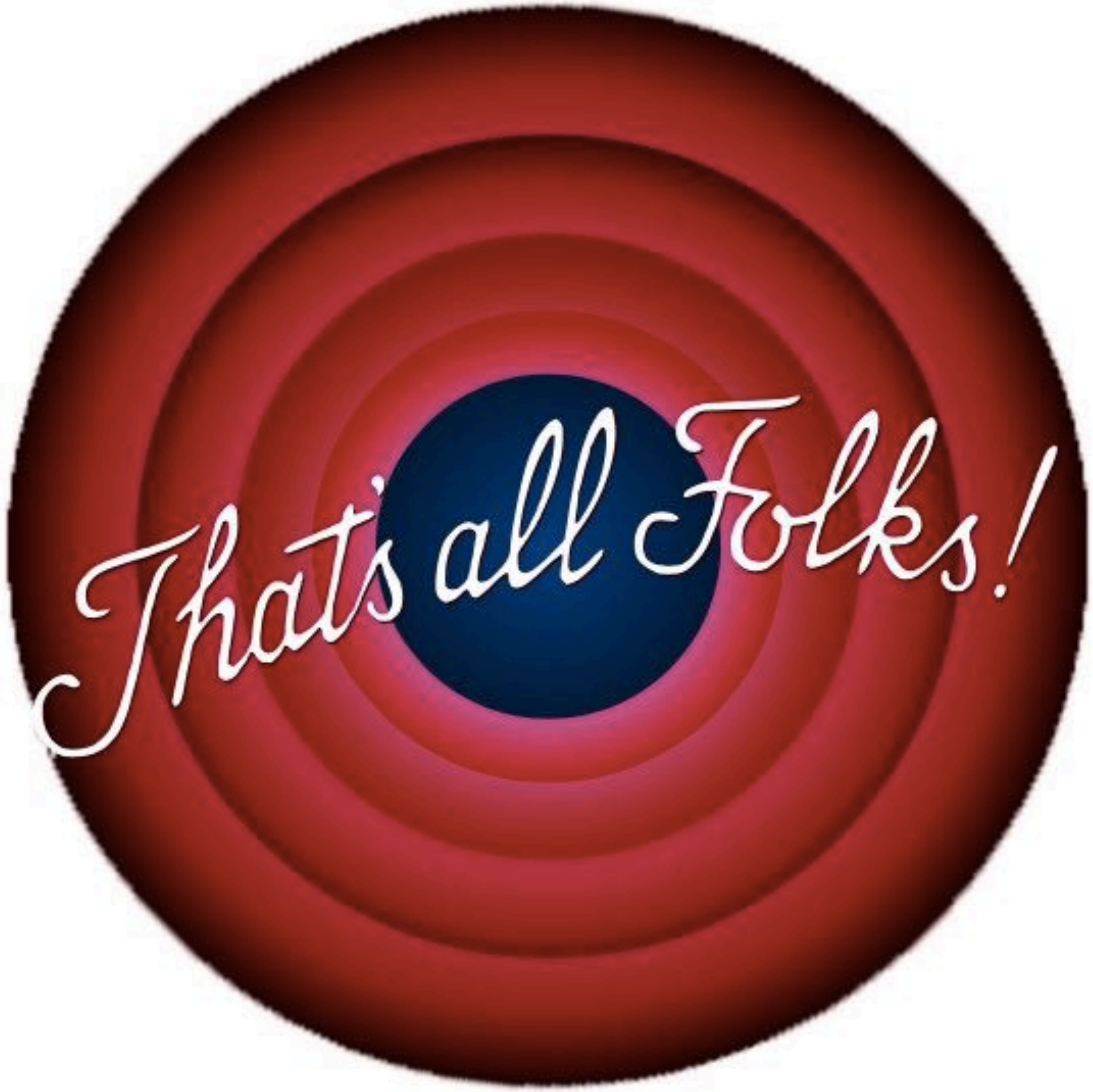
$$(\hat{\mathscr{X}}_i)_{[n]}\left(\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \quad \cdots \quad \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \quad \cdots \quad \odot \mathbf{U}^{(N)}\right)$$

$$vec(\mathbf{U}^{(n)}) = (\mathbf{\Phi}^{\top}\mathbf{\Phi} + \lambda \mathbf{I})^{-1}\mathbf{\Phi}^{\top}\mathbf{y}$$

@JeanKossaifi

Tensor Regression

- Practical example: Jupyter Notebook
- More reading:
  - W Guo, I. Kotsia, and I. Patras, *Tensor learning for regression,* IEEE Transactions on Image Processing, 21(2):816–827, Feb 2012
  - Qi Rose Yu and Yan Liu, *Learning from multiway data: Simple and efficient tensor regression,* CoRR, abs/1607.02535, 2016.
  - Guillaume Rabusseau and Hachem Kadri, *Low-rank regression with tensor responses,* In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (eds.), NIPS, pp. 1867–1875. 2016.



Any questions?

@JeanKossaifi
jean.kossaifi@gmail.com