

Low-Rank Tensor Regression

Jean Kossaifi



@JeanKossaifi
jean.kossaifi@gmail.com

Tensor Regression

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

$$\begin{aligned} \langle \hat{\mathcal{X}}, \hat{\mathcal{Y}} \rangle &= \sum_{i_0=0}^{I_0} \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=0}^{I_N} x_{i_0 i_1 \dots i_N} y_{i_0 i_1 \dots i_N} \\ &= \text{vec}(\hat{\mathcal{X}})^\top \text{vec}(\hat{\mathcal{Y}}) \end{aligned}$$

Tensor Regression

- Classical regression loss: $\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, \hat{\mathcal{W}} \rangle\|_F^2$
- Note that this is equivalent to linear regression as:

$$\langle \hat{\mathcal{X}}_i, \hat{\mathcal{W}} \rangle = \text{vec}(\hat{\mathcal{X}}_i)^\top \text{vec}(\hat{\mathcal{W}})$$

Tensor Regression

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

- Note that this is equivalent to linear regression as:

$$\langle \hat{\mathcal{X}}_i, \hat{\mathcal{W}} \rangle = \text{vec}(\hat{\mathcal{X}}_i)^\top \text{vec}(\hat{\mathcal{W}})$$

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

Tensor Regression

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

- Note that this is equivalent to linear regression as:

$$\langle \hat{\mathcal{X}}_i, \hat{\mathcal{W}} \rangle = \text{vec}(\hat{\mathcal{X}}_i)^\top \text{vec}(\hat{\mathcal{W}})$$

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

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

Tensor Regression

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

- Note that this is equivalent to linear regression as:

$$\langle \hat{\mathcal{X}}_i, \hat{\mathcal{W}} \rangle = \text{vec}(\hat{\mathcal{X}}_i)^\top \text{vec}(\hat{\mathcal{W}})$$

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

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

- Add regularisation on the weights

Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}|] \rangle\|_F^2$$

- Add regularisation on the weights

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

Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}] \rangle\|_F^2$$

- Add regularisation on the weights

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

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

- Separable regularisation term
- Closed-form solution

Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}|] \rangle\|_F^2 + \frac{\lambda}{2} \sum_{k=0}^N \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}|] \rangle\|_F^2 + \frac{\lambda}{2} \sum_{k=0}^N \|\mathbf{U}^{(\mathbf{k})}\|_F^2$$

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

$$(\mathbf{Trace}(\mathbf{AB}^\top) = \text{vec}(\mathbf{A})^\top \text{vec}(\mathbf{B}))$$

Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}] \rangle\|_F^2 + \frac{\lambda}{2} \sum_{k=0}^N \|\mathbf{U}^{(k)}\|_F^2$$

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

- Build Φ so that row k contains

$$(\hat{\mathcal{X}}_i)_{[n]} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \dots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \dots \odot \mathbf{U}^{(N)})$$

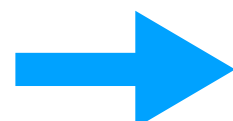
Tensor Regression

$$\frac{1}{2} \sum_{i=1}^N \|y_i - \langle \hat{\mathcal{X}}_i, [|\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}|] \rangle\|_F^2 + \frac{\lambda}{2} \sum_{k=0}^N \|\mathbf{U}^{(k)}\|_F^2$$

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

- Build Φ so that row k contains

$$(\hat{\mathcal{X}}_i)_{[n]} (\mathbf{U}^{(0)} \odot \mathbf{U}^{(1)} \odot \dots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \odot \dots \odot \mathbf{U}^{(N)})$$

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



Any questions?



@JeanKossaifi
jean.kossaifi@gmail.com