Comparison Table

Jiaxin Hu

May 19, 2021

Method	Model	# of features	non-Gaussian
STD (Ours)	$\mathbb{E}[\mathcal{Y}] = f(\mathcal{B} \times \{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}), \mathcal{B} = \mathcal{C} \times \{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\}$	3	$\surd$
Double-core[1]	$\mathbb{E}[\mathcal{Y}] = \mathcal{B}, \mathcal{B} = (\mathcal{C}_1 + \mathcal{C}_2) \times \{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\}$	0	$\surd$
GCP[4]	$\mathbb{E}[\mathcal{Y}] = f([\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!])$	0	$\surd$
CP-APR[3]	$\mathbb{E}[\mathcal{Y}] = f([\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!])$	0	Poi Only
CORALS[2]	$\mathbb{E}[\mathcal{Y}] = f([\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!])$	0	$\surd$
SupCP[6]	$\mathcal{Y} = [\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!] + \mathcal{E}, \boldsymbol{A}_1 = \boldsymbol{X}\boldsymbol{B} + \mathcal{E}'$	1	$\times$
mRRR[14]	$\mathcal{Y}_{ijk} \sim \exp \mathrm{fm}(\theta_{ijk}, \phi), \theta_{ijk} = f(\boldsymbol{X}\boldsymbol{B}), \mathrm{rank}(\boldsymbol{B}) = r$	1	$\surd$ and mixed
Envelope[5]	$\mathcal{Y} = \mathcal{B} \times_3 \boldsymbol{X} + \mathcal{E}, \mathcal{B} = \mathcal{C} \times \{\Gamma_1, \Gamma_2, \boldsymbol{I}_d\}, \mathrm{Cov}(\mathcal{E}) = \Sigma_1 \otimes \Sigma_2$	1	$\times$
GLSNet[12]	$\mathbb{E}[\mathcal{Y}] = f(\Theta + \mathcal{B} \times_3 \boldsymbol{X}), \mathrm{rank}(\Theta) = r, \Vert\mathcal{B}\Vert_0 = s$	1	$\surd$
STORE[7]	$\mathcal{Y} = \mathcal{B} \times_3 \boldsymbol{X} + \mathcal{E}, \mathcal{B} = [\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!], \Vert\boldsymbol{A}_k\Vert_0 \leq s_k$	1	$\times$
Han[10]	$y_i = \langle \mathcal{B}, \mathcal{X}_i \rangle + \epsilon, \mathcal{B} = \mathcal{C} \times \{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\}$	3	$\times$
Garvesh[8]	$y_i = \langle \mathcal{B}, \mathcal{X}_i \rangle + \epsilon, \mathcal{B}$ various structures	3	$\times$
STAR[9]	$\mathcal{Y}_{ijk} = \mathcal{T}(\mathcal{X}_i) + \epsilon, \mathcal{T}(\mathcal{X}_i) \approx \sum_m^M \langle \mathcal{B}_m, \mathcal{F}_m(\mathcal{X}_i) \rangle, \mathcal{B}_m$ CP sparse	3	$\times$

Table 1: Comparison of different methods in model, the largest number of feature matrices which are able to be incorporated, and whether the model capacity to deal with non-Gaussian data. Here we consider the observation $\mathcal{Y}$, which may be unfolded to matrix or vector based on formula of the model. Let $\mathcal{X}, \boldsymbol{X}, \boldsymbol{X}_k$ denote the feature tensors and matrices, $\mathcal{B}, \mathcal{B}_m, \boldsymbol{B}, \Theta$ denote the regression coefficient tensors and matrices, $\mathcal{C}$ denote the core tensor and $\boldsymbol{M}_k, \Gamma_k$factor matrices of Tucker decomposition, respectively, $\boldsymbol{A}_k$ denote the factors of tensor CP decomposition, and $\mathcal{E}, \mathcal{E}', \epsilon$ denote the noise. Besides, $f(\cdot)$ denote the link function, $\Vert\cdot\Vert_0$ denote the number of non-zero elements in a tensor or matrix, $\mathcal{T}(\cdot)$ denote a non-parametric function, $\mathcal{C} \times \{\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3\}$ denote the tucker product between the core tensor and factor matrices, $[\![\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3]\!]$ denote the outer-product in CP decomposition, Σ_k denote the covariance matrices, $\exp \mathrm{fm}(\theta_{ijk}, \phi)$ denote the exponential family with natural parameter θ_{ijk} and dispersion parameter ϕ. The dimension of the tensors and matrices can be determined by the context.

Method	Sparsity	non-i.i.d. noise	Algo	Algo guarantee	Complexity	Error bound
STD (Ours)	$\times$	$\times$	Alter/HOSVD	$\surd$	$r^3 + 3pr$	$\surd$
Double-core[1]	$\times$	$\times$	ADMM	$\surd$	$r^3 + 3dr$	$\surd$
GCP[4]	$\times(\surd)$	$\times$	BFGS	$\times$	$3dR$	$\times$
CP-APR[3]	$\times$	$\times$	Alter, MM	$\surd$	$3dR$	$\times$
CORALS[2]	$\surd$	$\times$	ALS	$\times$	$(3dR)^*$	$\times$
SupCP[6]	$\times$	$\times$	EM	$\times$	$2dR + pR$	$\times$
mRRR[14]	$\times$	$\times$	Alter	$\surd$	$pr + d^2r$	$\surd$
Envelope[5]	$\times$	$\surd$	Alter	$\times$	$(r^2d + 2dr)^+$	$\surd$
GLSNet[12]	$\surd$	$\times$	Alter GD	$\surd$	$2dr + s$	$\surd$
STORE[7]	$\surd$	$\times$	Alter	$\surd$	$r\sum_k s_k$	$\surd$
Han[10]	$\times$	$\times$	PGD	$\surd$	$r^3 + 3pr$	$\surd$
Garvesh[8]	$\times(\surd)$	$\times$	GD	$\times$	$(d^3)^*$	$\surd$
STAR[9]	$\surd$	$\times$	Alter	$\times$	$(3Mdr)^*$	$\surd$

Table 2: Comparison of different methods in sparsity assumption, non-i.i.d. noise assumption, algorithm, algorithm guarantee, model complexity, and error bound of the estimations. The mark $\surd$ means the paper consists of the discussion to the corresponding topics while $\times$ means no discussion in the paper. Here we consider the special case with d-by-d-by-d observation $\mathcal{Y}$, d-by-p feature matrices $\boldsymbol{X}, \boldsymbol{X}_k$, and the feature tensor transformed by the matrices $\mathcal{X} \in \mathbb{P}^{p\times p\times p}$. Assume the Tucker structure tensors have rank (r, r, r) and CP structure tensors have rank R. The value s, s_k refer to the sparsity, i.e., the number of non-zero elements based on the model assumption. The mark $\times(\surd)$ means the purposed method can be extended with sparsity assumption, $(\cdot)^*$ implies the model has soft sparsity assumption through some sparsity regularizers and thus the model complexity is related to the tuning parameters, and $(\cdot)^+$ implies that the Envelope method consists of extra complexity to estimate the covariance.