Adversarially-Trained Nonnegative Matrix Factorization

Ting Cai — University of Wisconsin-Madison

Vincent Y. F. Tan — National University of Singapore

Cédric Févotte — CNRS

ICASSP (Singapore), May 2022

Nonnegative Matrix Factorization (Lee and Seung, 1999)

- Given a nonnegative data matrix $\mathbf{V} \in \mathbb{R}_+^{F \times N}$, approximate $\mathbf{V}$ as

$$\mathbf{V} \approx \mathbf{WH}$$

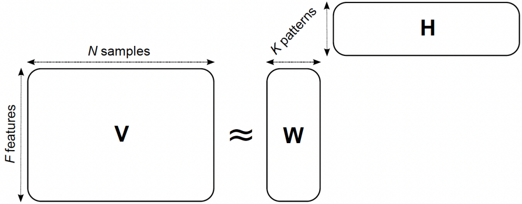
where $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ (basis) and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$ (coefficient) matrices.

Nonnegative Matrix Factorization (Lee and Seung, 1999)

- Given a nonnegative data matrix $\mathbf{V} \in \mathbb{R}_+^{F \times N}$, approximate $\mathbf{V}$ as

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

where $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ (basis) and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$ (coefficient) matrices.

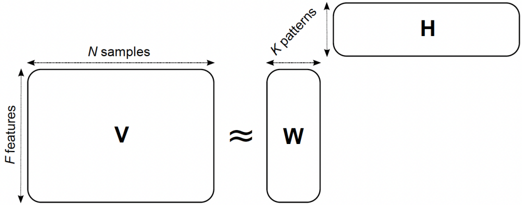


Nonnegative Matrix Factorization (Lee and Seung, 1999)

- Given a nonnegative data matrix $\mathbf{V} \in \mathbb{R}_{+}^{F \times N}$, approximate $\mathbf{V}$ as

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

where $\mathbf{W} \in \mathbb{R}_{+}^{F \times K}$ (basis) and $\mathbf{H} \in \mathbb{R}_{+}^{K \times N}$ (coefficient) matrices.



- One typically solves

$$\min_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} D(\mathbf{V} \mid \mathbf{W}\mathbf{H}) \quad \text{e.g.} \quad D(\mathbf{V} \mid \mathbf{W}\mathbf{H}) = \left\| \mathbf{V} - \mathbf{W}\mathbf{H} \right\|_{\mathrm{F}}^{2},$$

where $\mathbf{A} \geq \mathbf{0}$ means that all the entries of $\mathbf{A}$ are nonnegative.

Motivation and Main Contribution

- Adversarial training (Goodfellow et al., 2015; Madry et al., 2018; Tramèr et al., 2018)

$$\min_{\boldsymbol{\theta}\in\Theta} \underbrace{\max_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\leq\epsilon}}_{\text{perturbation of features}} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}\big(f_{\boldsymbol{\theta}}(x_i'), y_i\big).$$

Motivation and Main Contribution

- Adversarial training (Goodfellow et al., 2015; Madry et al., 2018; Tramèr et al., 2018)

$$\min_{\boldsymbol{\theta}\in\Theta} \underbrace{\max_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\leq\epsilon}}_{\text{perturbation of features}} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}\big(f_{\boldsymbol{\theta}}(x_i'), y_i\big).$$

- Improve the predictive performance of NMF using adversarial training for matrix completion tasks.

Motivation and Main Contribution

- Adversarial training (Goodfellow et al., 2015; Madry et al., 2018; Tramèr et al., 2018)

$$\min_{\boldsymbol{\theta}\in\Theta} \underbrace{\max_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\le\epsilon}}_{\text{perturbation of features}} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}\big(f_{\boldsymbol{\theta}}(x_i'), y_i\big).$$

- Improve the predictive performance of NMF using adversarial training for matrix completion tasks.
- Derive efficient algorithms for updating the adversary and $(\mathbf{W}, \mathbf{H})$.

Motivation and Main Contribution

- Adversarial training (Goodfellow et al., 2015; Madry et al., 2018; Tramèr et al., 2018)

$$\min_{\boldsymbol{\theta}\in\Theta} \underbrace{\max_{\mathbf{x}':\|\mathbf{x}-\mathbf{x}'\|\le\epsilon}}_{\text{perturbation of features}} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}\big(f_{\boldsymbol{\theta}}(x_i'), y_i\big).$$

- Improve the predictive performance of NMF using adversarial training for matrix completion tasks.
- Derive efficient algorithms for updating the adversary and $(\mathbf{W}, \mathbf{H})$.
- Demonstrate the superior predictive performance of adversarially-trained NMF or AT-NMF over other methods on matrix completion tasks for three benchmark datasets.

Formulation of AT-NMF

- Consider an adversary that adds an arbitrary matrix $\mathbf{R} \in \mathbb{R}_+^{F \times N}$ to $\mathbf{V}$ to maximize the divergence between $\mathbf{V}$ and $\mathbf{WH}$, AT-NMF is formulated as

$$\min_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \max_{\mathbf{R} \in \mathcal{R}} \left\| \mathbf{V} + \mathbf{R} - \mathbf{WH} \right\|_{\mathrm{F}}^2$$

where the constraint set

$$\mathcal{R} = \left\{ \mathbf{R} : \|\mathbf{R}\|_{\mathrm{F}}^2 \leq \epsilon, \mathbf{V} + \mathbf{R} \geq \mathbf{0} \right\}$$

Formulation of AT-NMF

- Consider an adversary that adds an arbitrary matrix $\mathbf{R} \in \mathbb{R}_+^{F \times N}$ to $\mathbf{V}$ to maximize the divergence between $\mathbf{V}$ and $\mathbf{WH}$, AT-NMF is formulated as

$$\min_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \max_{\mathbf{R} \in \mathcal{R}} \left\| \mathbf{V} + \mathbf{R} - \mathbf{WH} \right\|_{\mathrm{F}}^2$$

  where the constraint set

$$\mathcal{R} = \left\{ \mathbf{R} : \|\mathbf{R}\|_{\mathrm{F}}^2 \leq \epsilon, \mathbf{V} + \mathbf{R} \geq \mathbf{0} \right\}$$

- $\epsilon > 0$ is a constant indicating the adversary's power.

Formulation of AT-NMF

- Consider an adversary that adds an arbitrary matrix $\mathbf{R} \in \mathbb{R}_+^{F \times N}$ to $\mathbf{V}$ to maximize the divergence between $\mathbf{V}$ and $\mathbf{WH}$, AT-NMF is formulated as

$$\min_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \max_{\mathbf{R} \in \mathcal{R}} \left\| \mathbf{V} + \mathbf{R} - \mathbf{WH} \right\|_{\mathrm{F}}^2$$

where the constraint set

$$\mathcal{R} = \left\{ \mathbf{R} : \|\mathbf{R}\|_{\mathrm{F}}^2 \leq \epsilon, \mathbf{V} + \mathbf{R} \geq \mathbf{0} \right\}$$

- $\epsilon > 0$ is a constant indicating the adversary's power.

- To relax the problem, dualize the constraint $\|\mathbf{R}\|_{\mathrm{F}}^2 \leq \epsilon$ with Lagrange multiplier $\lambda > 0$, AT-NMF becomes

$$\min_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \max_{\mathbf{R}: \mathbf{V} + \mathbf{R} \geq \mathbf{0}} \left\| \mathbf{V} + \mathbf{R} - \mathbf{WH} \right\|_{\mathrm{F}}^2 - \lambda \left\| \mathbf{R} \right\|_{\mathrm{F}}^2.$$

AT-NMF Algorithm (Update of $\mathbf{R}$)

AT-NMF Algorithm (Update of $\mathbf{R}$)

- Let $\hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$, the inner maximization problem can be rewritten as a minimization problem as

$$\mathbf{R}^* = \underset{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}}{\arg\min} \; -\|\mathbf{V} + \mathbf{R} - \hat{\mathbf{V}}\|_{\mathrm{F}}^2 + \lambda \|\mathbf{R}\|_{\mathrm{F}}^2$$

AT-NMF Algorithm (Update of $\mathbf{R}$)

- Let $\hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$, the inner maximization problem can be rewritten as a minimization problem as

$$\mathbf{R}^* = \underset{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}}{\arg\min} \; -\|\mathbf{V} + \mathbf{R} - \hat{\mathbf{V}}\|_{\mathrm{F}}^2 + \lambda \, \|\mathbf{R}\|_{\mathrm{F}}^2$$

- Objective separates into FN independent terms

$$g(\mathbf{R}) = \sum_{f,n} \left[-(v_{fn} + r_{fn} - \hat{v}_{fn})^2 + \lambda \, r_{fn}^2 \right].$$

AT-NMF Algorithm (Update of $\mathbf{R}$)

- Let $\hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$, the inner maximization problem can be rewritten as a minimization problem as

$$\mathbf{R}^* = \arg\min_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} -\|\mathbf{V}+\mathbf{R}-\hat{\mathbf{V}}\|_{\mathrm{F}}^2 + \lambda\,\|\mathbf{R}\|_{\mathrm{F}}^2$$

- Objective separates into FN independent terms

$$g(\mathbf{R}) = \sum_{f,n}\left[-(v_{fn}+r_{fn}-\hat{v}_{fn})^2 + \lambda\, r_{fn}^2\right].$$

- Suffices to minimize each term inside over r_{fn}. By re-arranging:

$$\min_{r_{fn}:v_{fn}+r_{fn}\geq 0} (\lambda-1)r_{fn}^2 - 2r_{fn}(v_{fn}-\hat{v}_{fn})$$

- Can be solved in closed-form. For $\lambda \in [0,1]$, $r_{fn}=\infty$; for $\lambda > 1$,

$$\mathbf{R}^* = \max\left\{\frac{\mathbf{V}-\hat{\mathbf{V}}}{\lambda-1}, -\mathbf{V}\right\}$$

AT-NMF Algorithm (Update of $(\mathbf{W}, \mathbf{H})$)

AT-NMF Algorithm (Update of $(\mathbf{W}, \mathbf{H})$)

- After update of $\mathbf{R}$ to $\mathbf{R}^*$,

$$\underset{\mathbf{W},\mathbf{H}\geq\mathbf{0}}{\arg\min}\ \|\underbrace{\mathbf{V}+\mathbf{R}^*}_{=:\mathbf{U}}-\mathbf{WH}\|_{\mathrm{F}}^2 - \cancel{\lambda\|\mathbf{R}\|_{\mathrm{F}}^2}$$

fix $\mathbf{U} := \mathbf{V} + \mathbf{R}^*$.

AT-NMF Algorithm (Update of $(\mathbf{W}, \mathbf{H})$)

- After update of $\mathbf{R}$ to $\mathbf{R}^*$,

$$\arg\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \| \underbrace{\mathbf{V} + \mathbf{R}^*}_{=:\mathbf{U}} - \mathbf{W}\mathbf{H}\|_F^2 - \cancel{\lambda \|\mathbf{R}\|_F^2}$$

fix $\mathbf{U} := \mathbf{V} + \mathbf{R}^*$.

- Use majorization-minimization (MM) (Hunter and Lange, 2000) to update

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^\top \mathbf{U}}{\mathbf{W}^\top \mathbf{W}\mathbf{H}} \quad \text{and} \quad \mathbf{W} \leftarrow \mathbf{W} \cdot \frac{\mathbf{U}\mathbf{H}^\top}{\mathbf{W}\mathbf{H}\mathbf{H}^\top}$$

AT-NMF Algorithm (Update of $(\mathbf{W}, \mathbf{H})$)

- After update of $\mathbf{R}$ to $\mathbf{R}^*$,

$$\arg\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \|\underbrace{\mathbf{V}+\mathbf{R}^*}_{=:\mathbf{U}} - \mathbf{W}\mathbf{H}\|_F^2 - \cancel{\lambda\|\mathbf{R}\|_F^2}$$

fix $\mathbf{U} := \mathbf{V} + \mathbf{R}^*$.

- Use majorization-minimization (MM) (Hunter and Lange, 2000) to update

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^\top \mathbf{U}}{\mathbf{W}^\top \mathbf{W}\mathbf{H}} \quad \text{and} \quad \mathbf{W} \leftarrow \mathbf{W} \cdot \frac{\mathbf{U}\mathbf{H}^\top}{\mathbf{W}\mathbf{H}\mathbf{H}^\top}$$

- Initialization of $(\mathbf{W}, \mathbf{H})$
  - Sample each entry independently from Half-Normal distribution (with variance parameter $\gamma = 1$);
  - Run 5 standard MM steps on $\mathbf{V}$ to obtain $\mathbf{W}_{\text{init}}$ and $\mathbf{H}_{\text{init}}$.

AT-NMF Algorithm (Update of $(\mathbf{W}, \mathbf{H})$)

- After update of $\mathbf{R}$ to $\mathbf{R}^*$,

$$\operatorname*{arg\,min}_{\mathbf{W},\mathbf{H}\geq \mathbf{0}} \| \underbrace{\mathbf{V} + \mathbf{R}^*}_{=:\mathbf{U}} - \mathbf{W}\mathbf{H}\|_F^2 - \cancel{\lambda \|\mathbf{R}\|_F^2}$$

  fix $\mathbf{U} := \mathbf{V} + \mathbf{R}^*$.

- Use majorization-minimization (MM) (Hunter and Lange, 2000) to update

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^\top \mathbf{U}}{\mathbf{W}^\top \mathbf{W}\mathbf{H}} \quad \text{and} \quad \mathbf{W} \leftarrow \mathbf{W} \cdot \frac{\mathbf{U}\mathbf{H}^\top}{\mathbf{W}\mathbf{H}\mathbf{H}^\top}$$

- Initialization of $(\mathbf{W}, \mathbf{H})$
  - Sample each entry independently from Half-Normal distribution (with variance parameter $\gamma = 1$);
  - Run 5 standard MM steps on $\mathbf{V}$ to obtain $\mathbf{W}_{\text{init}}$ and $\mathbf{H}_{\text{init}}$.

- Termination (described in paper)

Synthetic Dataset: Setup

- $\alpha \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ denotes the fraction of held-out entries.
- $\Gamma \subset \{1, \cdots, F\} \times \{1, \cdots, N\}$ is the set of held-out entries of $\mathbf{V}$.
- $\hat{v}_{fn}$ is the prediction of v_{fn}.

Synthetic Dataset: Setup

- $\alpha \in \{0.1, 0.2, \cdots, 0.8, 0.9\}$ denotes the fraction of held-out entries.
- $\Gamma \subset \{1, \cdots, F\} \times \{1, \cdots, N\}$ is the set of held-out entries of $\mathbf{V}$.
- $\hat{v}_{fn}$ is the prediction of v_{fn}.

Synthetic Dataset: Metric and Results

- Our performance metric is the root mean-squared error (RMSE)

$$\text{RMSE} := \sqrt{\frac{1}{|\Gamma|} \sum_{(f,n)\in\Gamma} \left(v_{fn} - \hat{v}_{fn}\right)^2}$$

- Systhetic dataset of $F = 100$, $N = 50$, and $K = 5$.

Synthetic Dataset: Metric and Results

- Our performance metric is the root mean-squared error (RMSE)

$$\text{RMSE} := \sqrt{\frac{1}{|\Gamma|} \sum_{(f,n)\in\Gamma} \left(v_{fn} - \hat{v}_{fn}\right)^2}$$

- Systhetic dataset of $F = 100$, $N = 50$, and $K = 5$.
- RMSE of Synthetic dataset

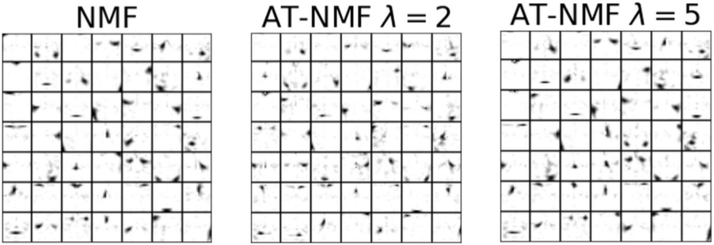
α	**NMF**	**ANMF**	**AT-NMF** (2)	**AT-NMF** (3)	**AT-NMF** (5)
0.3	5.37 ± 0.02	6.78 ± 0.17	5.41 ± 0.12	$\mathbf{5.11 \pm 0.03}$	5.20 ± 0.02
0.4	5.62 ± 0.03	6.92 ± 0.17	5.54 ± 0.08	$\mathbf{5.32 \pm 0.09}$	5.42 ± 0.04
0.5	6.41 ± 0.01	7.44 ± 0.09	6.27 ± 0.11	$\mathbf{6.05 \pm 0.03}$	6.18 ± 0.02
0.6	6.74 ± 0.02	7.61 ± 0.09	6.47 ± 0.07	$\mathbf{6.39 \pm 0.03}$	6.53 ± 0.02
0.7	7.30 ± 0.01	7.99 ± 0.06	7.02 ± 0.04	$\mathbf{6.94 \pm 0.01}$	7.10 ± 0.02
0.8	7.87 ± 0.01	8.30 ± 0.06	7.69 ± 0.04	$\mathbf{7.61 \pm 0.03}$	7.71 ± 0.00
0.9	8.45 ± 0.01	8.58 ± 0.06	8.44 ± 0.02	$\mathbf{8.34 \pm 0.02}$	8.35 ± 0.02

CBCL Face Dataset: Parts Learned

- $N = 2429$ facial images with $F = 361$ pixels.

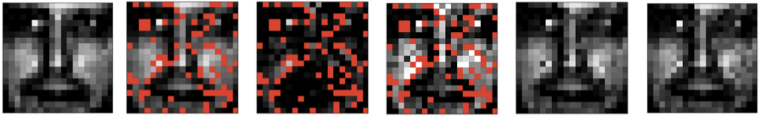
CBCL Face Dataset: Parts Learned

- $N = 2429$ facial images with $F = 361$ pixels.
- Parts learnt when $\alpha = 0.1$



CBCL Face Dataset: Image Restoration

Image Restoration by AT-NMF



(a) (b) (c) (d) (e) (f)

(a) Original Image $\mathbf{V}$;

(b) Masked training image $\mathbf{V}$;

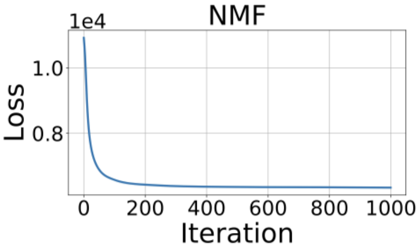
(c) Adversary's added-on masked image $\mathbf{R}^*$;

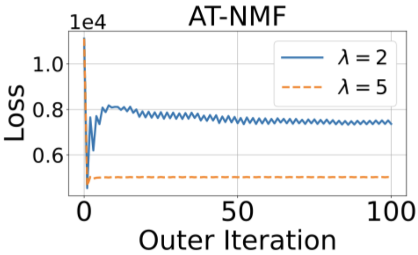
(d) AT-masked image $\mathbf{V} + \mathbf{R}^*$ [Features (eyes, nose, lower cheeks) become more distinctive];

(e) Restored image using AT-NMF with $\lambda = 2$;

(f) Restored image using NMF.

Training losses when $\alpha = 0.5$



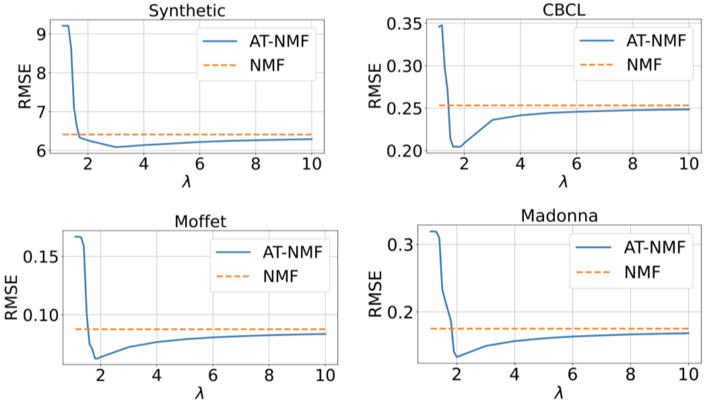


Hyperspectral Datasets

- It includes the Moffet (Jet Propulsion Lab) and Madonna (Sheeren et al., 2011) datasets with $F = 165$ and $F = 160$ respectively and $N = 2500$.

Hyperspectral Datasets

- It includes the Moffet (Jet Propulsion Lab) and Madonna (Sheeren et al., 2011) datasets with $F = 165$ and $F = 160$ respectively and $N = 2500$.
- Effect of λ on the RMSE



Conclusions and Future Work

- Formulation and algorithm for Adversarially-Trained NMF:

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}\in\mathcal{R}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^{2}$$

where $\mathcal{R}=\{\mathbf{R}:\|\mathbf{R}\|_{\mathrm{F}}^{2}\leq\epsilon,\mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$ or

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^{2}-\lambda\left\|\mathbf{R}\right\|_{\mathrm{F}}^{2}.$$

Conclusions and Future Work

- Formulation and algorithm for Adversarially-Trained NMF:

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}\in\mathcal{R}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2$$

where $\mathcal{R}=\{\mathbf{R}:\|\mathbf{R}\|_{\mathrm{F}}^2\leq\epsilon, \mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$ or

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2-\lambda\left\|\mathbf{R}\right\|_{\mathrm{F}}^2.$$

- Other divergence measures beyond the Frobenius norm , e.g., β-divergence (Févotte et al., 2009).

Conclusions and Future Work

- Formulation and algorithm for Adversarially-Trained NMF:

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}\in\mathcal{R}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2$$

  where $\mathcal{R} = \{\mathbf{R} : \|\mathbf{R}\|_{\mathrm{F}}^2 \leq \epsilon, \mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$ or

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2 - \lambda\left\|\mathbf{R}\right\|_{\mathrm{F}}^2.$$

- Other divergence measures beyond the Frobenius norm , e.g., β-divergence (Févotte et al., 2009).
- Different bounded sets $\mathcal{R}$, e.g., $\{\mathbf{R} : \|\mathbf{R}\|_{p,q} \leq \epsilon, \mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$.

Conclusions and Future Work

- Formulation and algorithm for Adversarially-Trained NMF:

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}\in\mathcal{R}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2$$

  where $\mathcal{R}=\{\mathbf{R}:\|\mathbf{R}\|_{\mathrm{F}}^2\leq\epsilon,\mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$ or

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} \max_{\mathbf{R}:\mathbf{V}+\mathbf{R}\geq\mathbf{0}} \left\|\mathbf{V}+\mathbf{R}-\mathbf{W}\mathbf{H}\right\|_{\mathrm{F}}^2-\lambda\left\|\mathbf{R}\right\|_{\mathrm{F}}^2.$$

- Other divergence measures beyond the Frobenius norm , e.g., β-divergence (Févotte et al., 2009).
- Different bounded sets $\mathcal{R}$, e.g., $\{\mathbf{R}:\|\mathbf{R}\|_{p,q}\leq\epsilon,\mathbf{V}+\mathbf{R}\geq\mathbf{0}\}$.
- Online NMF (Lefèvre et al., 2011; Mairal, 2015)?

References I

C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural computation*, 21(3):793–830, 2009.

I. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. In *Proc. of the 3rd International Conference on Learning Representations*, 2015. URL `http://arxiv.org/abs/1412.6572`.

D. R. Hunter and K. Lange. Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, 9(1):60–77, 2000.

Jet Propulsion Lab. Aviris free data. California Inst. Technol., Pasadena, CA, 2006. [Online]. Available: `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`.

D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

A. Lefèvre, F. Bach, and C. Févotte. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY, Oct. 2011. URL `https://www.irit.fr/~Cedric.Fevotte/publications/proceedings/waspaa11.pdf`.

A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. In *Proc. of the 6th International Conference on Learning Representations*, 2018.

J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM Journal on Optimization*, 25(2):829–855, 2015.

References II

D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon. Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS), Vancouver, Canada*, pages 3672–3675, July 2011.

F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel. Ensemble adversarial training: Attacks and defenses. In *Proc. of the 6th International Conference on Learning Representations*, 2018.