# Neuron-Enhanced AutoEncoder Matrix Completion and Collaborative Filtering: Theory and Practice

Jicong Fan[1,2] Rui Chen[1,2] Zhao Zhang[3] Chris H.Q. Ding[1]

[1]The Chinese University of Hong Kong, Shenzhen, China
[2]Shenzhen Research Institute of Big Data, Shenzhen, China
[3]Hefei University of Technology, China

- Propose AEMC-NE for matrix completion and collaborative filtering.
- Provide generalization error bounds regarding different missing mechanisms.

# Motivation

- Element-wise nonlinearity widely exists in real data
  - The observed variable $\boldsymbol{x} \in \mathbb{R}^d$ is given by $x_i = f(z_i)$, $i = 1, \ldots, d$, where $\boldsymbol{z} \in \mathbb{R}^d$ is the unobserved variable and $f : \mathbb{R} \to \mathbb{R}$ is a nonlinear function (usually unknown).
- Examples
  - In imaging science, the intensity of pixels are nonlinear responses of photoelectric element to the spectrum.
  - In chemical engineering, many sensors have nonlinear responses.
  - In biomedical engineering, the dose-responses are often nonlinear curves.
  - In recommendations systems, the ratings may be nonlinear responses to some latent values, according to the studies on response curves in psychology.

# Proposed Model

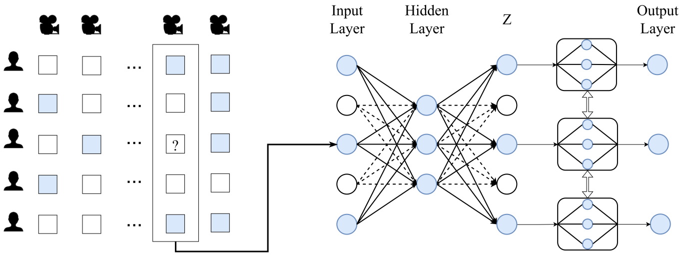
AEMC-NE is composed of two different neural networks.

- $g$ is an autoencoder to reconstruct the incomplete rating matrix $\tilde{\boldsymbol{X}}$

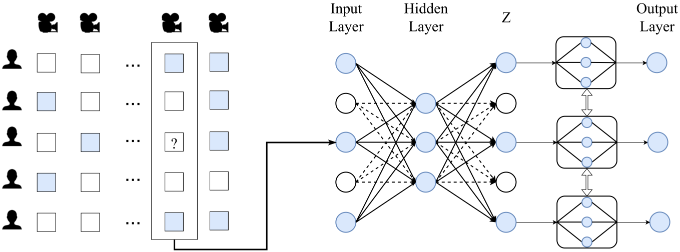
$$g_W(\tilde{\boldsymbol{X}}) = \boldsymbol{W}_{L_W}(\sigma_W(\boldsymbol{W}_{L_W-1}\sigma_W(\cdots\sigma_W(\boldsymbol{W}_1\tilde{\boldsymbol{X}})\cdots)))$$

- $h$ is an element-wise neural network to learn an activation function for the output of $g$

$$h_\Theta(z) = \boldsymbol{\Theta}_{L_\Theta}(\sigma_\Theta(\boldsymbol{\Theta}_{L_\Theta-1}\sigma_\Theta(\cdots\sigma_\Theta(\boldsymbol{\Theta}_1 z)\cdots)))$$



# Optimization




$$\underset{W,\Theta}{\text{minimize}} \left\| \boldsymbol{S} \odot \left( \tilde{\boldsymbol{X}} - h_\Theta\left(g_W(\tilde{\boldsymbol{X}})\right)\right) \right\|_F^2 + \lambda_W \sum_{l=1}^{L_W} \|\boldsymbol{W}_l\|_F^2 + \lambda_\Theta \sum_{l=1}^{L_\Theta} \|\boldsymbol{\Theta}_l\|_F^2$$

# Experiments on Synthetic Data

$\boldsymbol{X} = \boldsymbol{W}_3 Tanh(\boldsymbol{W}_2 Tanh(\boldsymbol{W}_1 \boldsymbol{Z})), \boldsymbol{X} = (Cosine(\boldsymbol{X}) + \boldsymbol{X})$, where $\boldsymbol{W}_1 \in \mathbb{R}^{50\times 10}$, $\boldsymbol{W}_2 \in \mathbb{R}^{100\times 50}$, $\boldsymbol{W}_3 \in \mathbb{R}^{300\times 100}$, and $\boldsymbol{Z} \in \mathbb{R}^{10\times 3000}$.

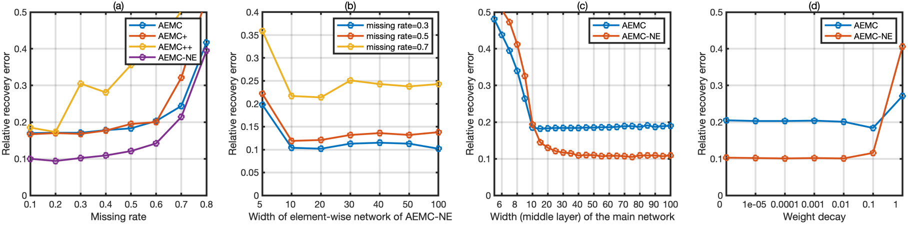



Figure: Recovery performance on synthetic data ($300 \times 3000$) in the setting of **MCAR**. **(a)** Influence of missing rate. Network structure: AEMC 300-100-30-100-300; AEMC+ 300-100-30-100-100-300; AEMC++ 300-100-30-100-100-100-300; AEMC-NE 300-100-30-100-300 for the main network and $1 - w - w - 1$ for the element-wise network, where $w = 20$. **(b)** AEMC-NE with different width ($w$) of the hidden layers in the element-wise network. **(c)** Influence of the width (middle layer) of the main network. The missing rate is 0.5 and $w = 20$. **(d)** Influence of the weight decay ($\lambda_W = \lambda_\Theta$). The missing rate is 0.5 and $w = 20$.

# Experiments on Synthetic Data

$\boldsymbol{X} = \boldsymbol{W}_3 Tanh(\boldsymbol{W}_2 Tanh(\boldsymbol{W}_1 \boldsymbol{Z}))$, $\boldsymbol{X} = (Cosine(\boldsymbol{X}) + \boldsymbol{X})$, where $\boldsymbol{W}_1 \in \mathbb{R}^{50\times 10}$, $\boldsymbol{W}_2 \in \mathbb{R}^{100\times 50}$, $\boldsymbol{W}_3 \in \mathbb{R}^{300\times 100}$, and $\boldsymbol{Z} \in \mathbb{R}^{10\times 3000}$.

| Missing rate | 0.2 | 0.3 | 0.4 |
|---|---|---|---|
| AEMC | $0.170_{\pm 0.002}$ | $0.171_{\pm 0.001}$ | $0.180_{\pm 0.003}$ |
| AEMC-NE | $0.101_{\pm 0.002}$ | $0.106_{\pm 0.003}$ | $0.109_{\pm 0.004}$ |
| Missing rate | 0.5 | 0.6 | 0.7 |
| AEMC | $0.187_{\pm 0.002}$ | $0.211_{\pm 0.003}$ | $0.268_{\pm 0.004}$ |
| AEMC-NE | $0.124_{\pm 0.004}$ | $0.153_{\pm 0.005}$ | $0.243_{\pm 0.006}$ |

Table: Rcovery performance on synthetic data in the setting of **MNAR**

# RMSE Results on Three MovieLens Datasets

| **Model** | **ML-100k** | **ML-1M** | **ML-10M** |
|---|---|---|---|
| BiasMF | 0.911 | 0.845 | 0.803 |
| LLORMA | 0.8881 | 0.833 | 0.782 |
| GC-MC | 0.905 | 0.832 | 0.777 |
| AutoSVD++ | 0.904 | 0.848 | - |
| AutoSVD | 0.901 | 0.86 | - |
| CF-NADE | - | 0.829 | 0.771 |
| DMF+ | 0.8889 | 0.8321 | - |
| IMC-GAE | 0.897 | 0.829 | - |
| GHRS | 0.8887 | 0.833 | 0.782 |
| AEMC (AutoRec) | - | 0.831± 0.003 | 0.782± 0.003 |
| AEMC (ReLU) | 0.8807 ± 0.0092 | 0.8276± 0.0025 | 0.7851 ± 0.0029 |
| **AEMC-NE** (ours) | **0.8767** ± 0.0089 | **0.8248** ± 0.0024 | **0.7723** ± 0.0025 |

# The End

More numerical results as well as the generalization bounds can be found in our paper.

Thanks for your attention!