EE269
Signal Processing for Machine Learning

Lecture 17

Instructor : Mert Pilanci

Stanford University

November 18, 2020

Nonnegative Matrix Factorization (NMF)

X : $n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} ||X - WZ||_F^2$$

- non-negative data
- pixel values, energies, counts are non-negative valued

Nonnegative Matrix Factorization

- PCA and Fisher's LDA basis vectors are hard to interpret

Nonnegative Matrix Factorization

X : $n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} ||X - WZ||_F^2$$

- No closed form solution!
- Solution is not unique: $X = WDD^{-1}Z = WZ$

Nonnegative Matrix Factorization

X : $n \times d$ data matrix

$$\min_{W \geq 0, Z \geq 0} ||X - WZ||_F^2$$

- No closed form solution!
- Solution is not unique: $X = WDD^{-1}Z = WZ$
- Heuristic algorithm:

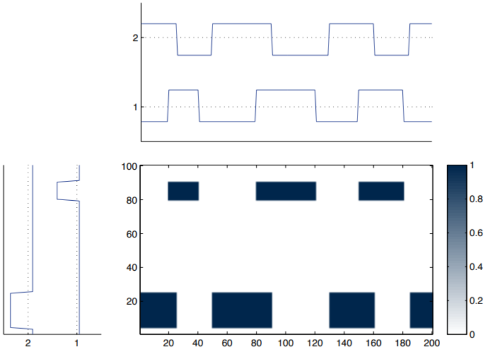
  $W \leftarrow \arg\min_W \; ||X - WZ||_F = XZ^\dagger, \quad W \leftarrow (W)_+$

  $Z \leftarrow \arg\min_Z \; ||X - WZ||_F = W^\dagger X, \quad Z \leftarrow (Z)_+$

  iterate

Nonnegative Matrix Factorization

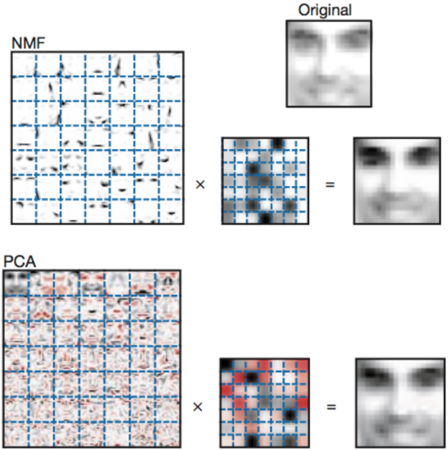
$$\min_{W \geq 0, Z \geq 0} ||X - WZ||_F^2 \qquad \text{implies} \qquad X \approx WZ$$



- Spectrogram model: columns of W contain frequencies (e.g. musical notes), rows of Z contain time information

Nonnegative Matrix Factorization for Images

$$\min_{W \geq 0, Z \geq 0} ||X - WZ||_F^2 \qquad \text{implies} \qquad X \approx WZ$$



Semi-Nonnegative Matrix Factorization (Semi-NMF)

Y : $n \times d$ data matrix

$$\min_{W \geq 0, Z} ||WZ - Y||_F^2$$

- Second component Z is not constrained

Semi-NMF and ReLU Neural Networks

- Semi-NMF

$$\min_{W \geq 0, Z} ||WZ - Y||_F^2$$

- Two-Layer ReLU neural network

$$\min_{W_1 W_2} ||(XW_1)_+ W_2 - Y||_F^2$$

- $(u)_+$ is the ReLU activation

[1]Need $n \leq d$ for this condition to hold

Semi-NMF and ReLU Neural Networks

- Semi-NMF

$$\min_{W \geq 0, Z} ||WZ - Y||_F^2$$

- Two-Layer ReLU neural network

$$\min_{W_1 W_2} ||(XW_1)_+ W_2 - Y||_F^2$$

- $(u)_+$ is the ReLU activation
- Suppose that the data matrix is whitened and we have $XX^T = I$. [1]
- substitute $W_1 \leftarrow X^T W_1$

[1] Need $n \leq d$ for this condition to hold

Semi-NMF and ReLU Neural Networks

- Semi-NMF

$$\min_{W \geq 0, Z} ||WZ - Y||_F^2$$

- Two-Layer ReLU neural network for whitened data $XX^T = I$

$$\min_{W_1 W_2} ||(W_1)_+ W_2 - Y||_F^2 = \min_{W_1 \geq 0} ||W_1 W_2 - Y||_F^2$$

- identical to the Semi-NMF model

Deep NMF and Semi-NMF

$$\min_{W_1 \geq 0, W_2 \geq 0, \ldots, W_{L-1} \geq 0, W_L} ||W_1 W_2 \ldots W_L - Y||_F^2$$

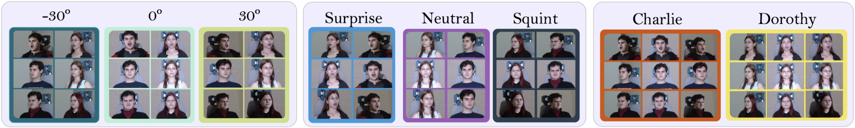
- Works well for speech, signal and image separation/clustering

Deep NMF and Semi-NMF

$$\min_{W_1 \geq 0, W_2 \geq 0, \ldots, W_{L-1} \geq 0, W_L} ||W_1 W_2 \ldots W_L - Y||_F^2$$

- Works well for speech, signal and image separation/clustering

A Deep Semi-NMF Model for Learning Hidden Representations Trigeorgis et al.



Clustering

$$\min \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - c_j\|^2 \; = \; \|X - CB^T\|_F^2$$

$C = [c_1, ..., c_k]$ is the centroid matrix

B denotes the clustering assignment

$B_{ij} = 1$ if i-th observation is assigned to the j-th cluster center c_j