

Series 5, March 30-31, 2017 (Non-Negative Matrix Factorization)

Problem 1 (Implementing NMF for Image Analysis):

Recall the non-negative matrix factorization (NMF) definition with quadratic cost, given as

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \quad & J(\mathbf{U}, \mathbf{V}) = \frac{1}{2} \|\mathbf{X} - \mathbf{U}^\top \mathbf{V}\|_F^2. \\ \text{s.t.} \quad & u_{zi}, v_{zj} \geq 0 \quad (\forall i, j, z) \quad (\text{non-negativity}). \end{aligned}$$

Implement the alternating least squares algorithm for the same face dataset as in Exercise 2. Use projection to non-negative entries, applied after every iteration of the algorithm.

Setup:

- As in Exercise 2, download the face images dataset, as well as the provided IPython notebook template from the lecture's github repository

https://github.com/dalab/lecture_cil_public/tree/master/exercises/ex5

- Build a matrix collecting all images as its rows (but this time do *not* normalize the images by their mean), as given in the template code.
- Initialize \mathbf{U} and \mathbf{V} randomly with non-negative entries.

Run the algorithm for say 20 iterations of alternating least squares, and finally visualize the $K = 3$ components (rows of the matrix \mathbf{V}). What is the interpretation of these images? What happens if you perform more iterations and/or increase the rank K ? Comment on the differences to Exercise 2, where we have used PCA instead for the same task.

Problem 2 (PLSA):

Recall the log-likelihood derived in class,

$$\mathcal{L}(\mathbf{U}, \mathbf{V}) = \sum_{i,j} x_{ij} \log p(w_j | d_i) = \sum_{(i,j) \in \mathcal{X}} \log \sum_{z=1}^K \underbrace{p(w_j | z)}_{=: v_{zj}} \underbrace{p(z | d_i)}_{=: u_{zi}},$$

as well as the variational lower bound

$$Q(\mathbf{U}, \mathbf{V}) := \sum_{i,j} x_{ij} \sum_{z=1}^K q_{zij} [\log u_{zi} + \log v_{zj} - \log q_{zij}] \leq \sum_{i,j} x_{ij} \log \sum_{z=1}^K q_{zij} \frac{u_{zi} v_{zj}}{q_{zij}}$$

1) Construct the Lagrangian of $Q(\mathbf{U}, \mathbf{V})$ by considering the two following constraints:

- $u_{zi} \geq 0$ such that $\sum_z u_{zi} = 1 \quad (\forall i)$
- $v_{zj} \geq 0$ such that $\sum_j v_{zj} = 1 \quad (\forall z)$

2) Show that the optimal parameters can be derived by optimizing the Lagrangian you derived in step 1, leading to the following expressions:

$$u_{zi} = \frac{\sum_j x_{ij} q_{zij}}{\sum_j x_{ij}}, \quad v_{zj} = \frac{\sum_i x_{ij} q_{zij}}{\sum_{i,l} x_{il} q_{zil}},$$