Exercises
Computational Intelligence Lab
SS 2017

Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 5 Solutions (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 entires.

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.

Answer:

The images corresponding to the rows of the matrix $\mathbf{V}$ should correspond to parts of faces. For a larger input dataset of original faces, we'd expect even more clear separation of part features.

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

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

1) Construct the Lagrangian of $Q(\mathbf{U},\mathbf{V})$ by considering the two following constraints:
   - $u_{zi} \geq 0$ such that $\sum_z u_{zi} = 1$ $(\forall i)$
   - $v_{zj} \geq 0$ such that $\sum_j v_{zj} = 1$ $(\forall z)$

Answer:

Let α, β to be the multipliers for u, v, respectively, then the Lagrangian function is (arguments except for α, β skipped for simplicity):

$$\mathcal{L}_a(\alpha, \beta) = \sum_{i,j} x_{ij} \sum_{z=1}^{k} q_{zij} \left[\log u_{zi} + \log v_{zj} - \log q_{zij}\right] + \sum_i \alpha_i (\sum_z u_{zi} - 1) + \sum_z \beta_z (\sum_j v_{zj} - 1) \quad (1)$$

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

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

Answer:

From the Expectation step, one can get the update rule of q_{zij} (see lecture slides), and $\sum_z q_{zij} = 1, \forall i, j$.

For u_{zi}:

$$\frac{\partial \mathcal{L}_a(\alpha, \beta)}{\partial u_{zi}} = \sum_j x_{ij} q_{zij} / u_{zi} + \alpha_i \stackrel{!}{=} 0 \quad (2)$$

one can get

$$u_{zi} = \frac{-\sum_j x_{ij} q_{zij}}{\alpha_i} \quad (3)$$

Take summation over z over both sides of the above equation, one get

$$1 = \sum_z u_{zi} = \frac{-\sum_j x_{ij} \sum_z q_{zij}}{\alpha_i} = \frac{-\sum_j x_{ij}}{\alpha_i}$$

So $\alpha_i = -\sum_j x_{ij}$, plug which into (3) one can get $u_{zi} = \frac{\sum_j x_{ij} q_{zij}}{\sum_j x_{ij}}$.

For v_{zj},

$$\frac{\partial \mathcal{L}_a(\alpha, \beta)}{\partial v_{zj}} = \sum_i x_{ij} q_{zij} / v_{zj} + \beta_z \stackrel{!}{=} 0 \quad (4)$$

one can get

$$v_{zj} = \frac{-\sum_i x_{ij} q_{zij}}{\beta_z} \quad (5)$$

Take summation over j on both sides of (5), it reads,

$$1 = \sum_j v_{zj} = \frac{-\sum_{ij} x_{ij} q_{zij}}{\beta_z}$$

so $\beta_z = -\sum_{ij} x_{ij} q_{zij}$, plug which into (5), we can get,

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