Exercises
Computational Intelligence Lab
SS 2018

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

Series 6, Solutions (Word Embeddings and Text Classification)

Solution 1 (Theory of Word Embeddings):

Recall the GloVe objective that consists in weighted least squares fit of log-counts, written as

$$\mathcal{H}(\theta; \mathbf{N}) = \sum_{i,j:\, n_{ij}>0} f(n_{ij}) \Big(\underbrace{\log n_{ij}}_{\text{target}} - \underbrace{\langle \mathbf{x}_i, \mathbf{y}_j \rangle}_{\text{model}} \Big)^2,$$

where $\langle \mathbf{x}_i, \mathbf{y}_j \rangle = \log \tilde{p}_\theta(w_i | w_j)$. Note that we ignore here the bias terms for simplicity.

1) Assume $f(\cdot) \equiv 1$ for all arguments, and write $m_{ij} := \log n_{ij}$.

   a) According to the chain rule,

$$\frac{\partial \mathcal{H}}{\partial \mathbf{x}_i} = \sum_{j:\, n_{ij}>0} 2(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{y}_j, \tag{1}$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{y}_j} = \sum_{i:\, n_{ij}>0} 2(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{x}_i. \tag{2}$$

   b) Let $\mathcal{H}_{ij} := f(n_{ij})(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij})^2$, so the stochastic gradient is:

$$\frac{\partial \mathcal{H}_{ij}}{\partial \mathbf{x}_i} = 2(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{y}_j, \tag{3}$$

$$\frac{\partial \mathcal{H}_{ij}}{\partial \mathbf{y}_j} = 2(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{x}_i. \tag{4}$$

2) Plug in the specific weight function:

$$\mathcal{H} = \sum_{i,j:\, n_{ij}>0} (m_{ij} - \langle \mathbf{x}_i, \mathbf{y}_j \rangle)^2 = \sum_{i,j:\, n_{ij}>0} \big(m_{ij} - (\mathbf{X}^\top \mathbf{Y})_{ij}\big)^2 \tag{5}$$

   the minimization of which w.r.t. $\mathbf{X}, \mathbf{Y}$ is a problem of matrix completion.

3)

$$\frac{\partial \mathcal{H}}{\partial \mathbf{x}_i} = \sum_{j:\, n_{ij}>0} 2 f(n_{ij})(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{y}_j \tag{6}$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{y}_j} = \sum_{i:\, n_{ij}>0} 2 f(n_{ij})(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{x}_i \tag{7}$$

$$\frac{\partial \mathcal{H}_{ij}}{\partial \mathbf{x}_i} = 2 f(n_{ij})(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{y}_j \tag{8}$$

$$\frac{\partial \mathcal{H}_{ij}}{\partial \mathbf{y}_j} = 2 f(n_{ij})(\langle \mathbf{x}_i, \mathbf{y}_j \rangle - m_{ij}) \mathbf{x}_i \tag{9}$$

Problem 3 (GloVe Implementation):

The solution file is in the public git repository:

`https://github.com/dalab/lecture_cil_public/tree/master/exercises/ex6`

Hint for Step 3): you can use PCA provided by scikit-learn, e.g. http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_3d.html or any other library.