Machine Learning Exercise Sheet 11

Dimensionality Reduction & Matrix Factorization

Homework

t-SNE

Problem 1: The similarity in the low dimensional space is defined as:

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_k \sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}}$$

The objective is to obtain a low-dimensional projection capturing the similarity structure of the high-dimensional data. This is achieved via optimizing the Kullback-Leibler divergence

$$C = KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Please derive the gradient $\frac{\partial C}{\partial y_s}$ for t-SNE for the coordinate y_i in the low dimensional space. Please note that this gradient can be used to update y_s with first-order methods.

Autoencoders

Problem 2: We train a linear autoencoder to D-dimensional data. The autoencoder has a single K-dimensional hidden layer, there are no biases, and all activation functions are identity ($\sigma(x) = x$).

- Why is it usually impossible to get zero reconstruction error in this setting if $K < D$?
- Under which conditions is this possible?

Coding Exercise

Problem 3: Download the notebook `exercise_11_notebook.ipynb` and `exercise_11_matrix_factorization_ratings.npy` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to PDF and append it to your other solutions before uploading.

Upload a single PDF file with your homework solution to Moodle by 03.02.2021, 11:59pm CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.