Machine Learning Exercise Sheet 11

Dimensionality Reduction & Matrix Factorization

In-class Exercise

There is no in-class exercise this week.

Homework

t-SNE

Problem 1: Figure 1 shows a scatter plot of your two-dimensional data ($N = 13$ instances). You want to apply a non-linear dimensionality reduction technique based on neighbor graphs (e.g. T-SNE or UMAP). As a first step you compute the $N \times N$, weighted adjacency matrix representing the neighbor graph. Assume that the weights are computed as

$$p_{j|i} = \frac{\exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 / 2\sigma^2\right)}{\sum_{k \neq i} \exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_k\right\|^2 / 2\sigma^2\right)}$$

where $\boldsymbol{x}_i \in \mathbb{R}^2$ and you set $p_{i|i} = 0$. Finally, you obtain the similarity between instances i and j with $p_{ij} = \frac{p_{i|j} + p_{j|i}}{2}$.

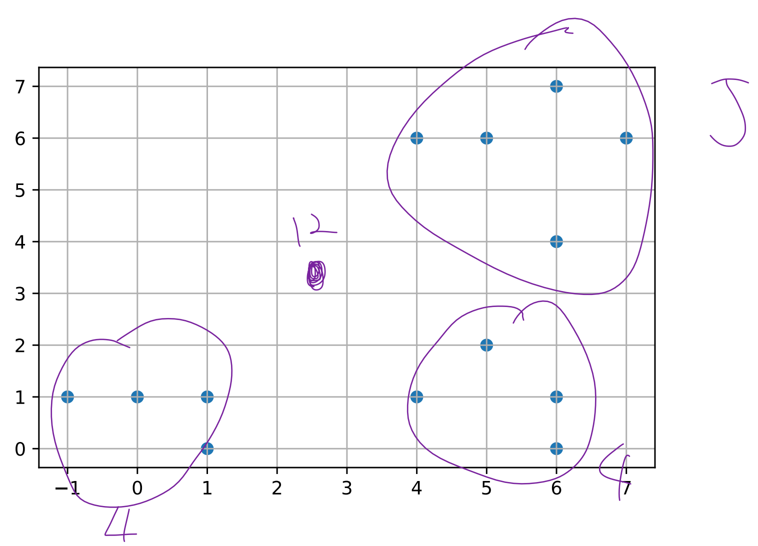


Figure 1: Scatter plot of the data

Which of the following neighbor graph plots (pixel in position i, j shows the value of p_{ij}) corresponds to the given dataset and the stated formula for $\sigma = 2$? What is your answer for $\sigma = 5$? *Justify your answers!*

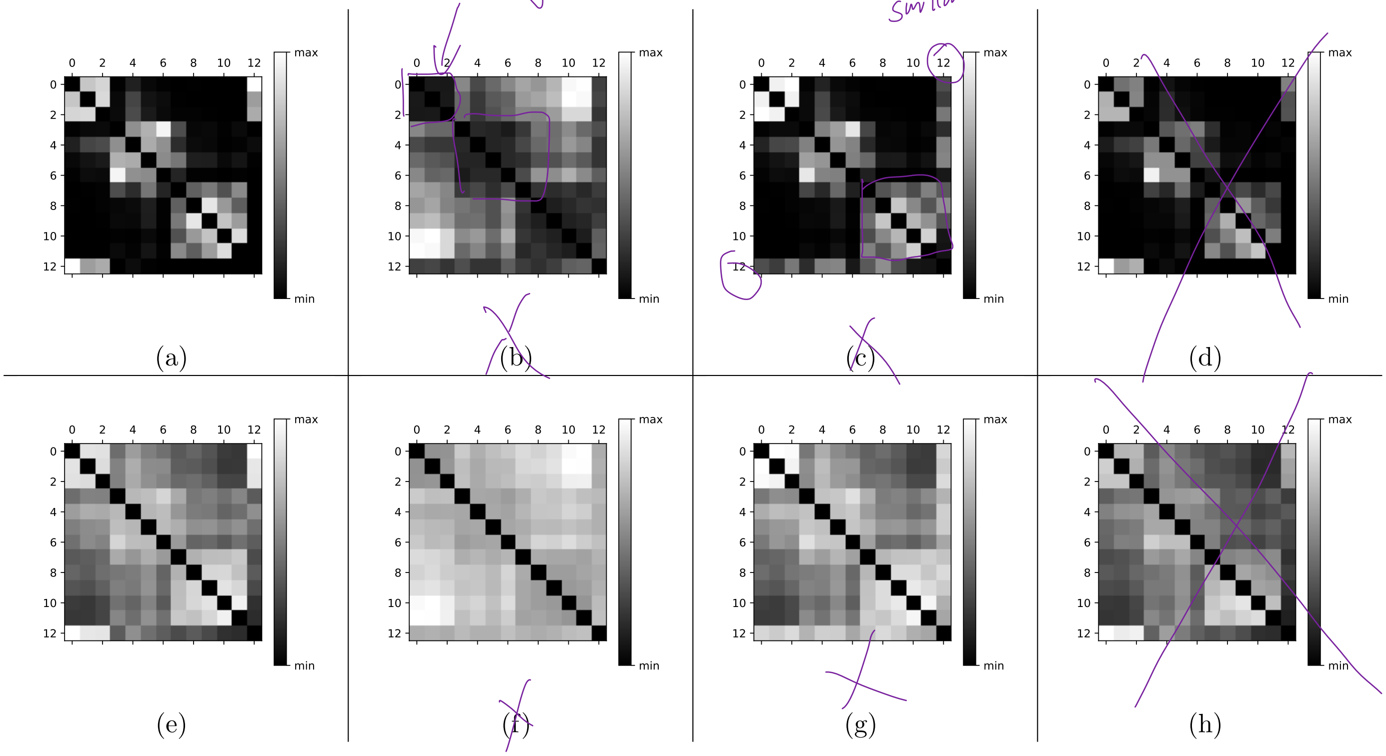
symmetric Matrix

wrong

similar to each other

Low σ ⇒

High σ ⇒



(a) (b) (c) (d)

(e) (f) (g) (h)

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?

$x \in \mathbb{R}^{N \times D}$

$f_{enc}(x, W_1) = x W_1$

$W_1 \in \mathbb{R}^{D \times K}$

$f_{dec}(z, W_2) = z W_2$

$W_2 \in \mathbb{R}^{K \times D}$

If rank(X) = K

K = D ?

$K < D$ W_2 Here K-dim to D-dim will cause some dimensional problem

Coding Exercise

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