MATH 250: Mathematical Data Visualization
Spring 2025

Homework 6: Manifold learning and Linear Discriminant Analysis

For all questions below, you should show all work needed to reach your answer. You may collaborate with your classmates and consult external resources, but you should write and submit your own answer. **Any classmates with whom you collaborate should be credited at the top of your submission. Similarly, if you consult any external references, you should cite them clearly and explicitly.**

1. Stochastic neighbor embedding balances attraction (between nearby points in the original feature space) and repulsion (between nearby points in the embedding space). First, we define a graph whose nodes are the original high dimensional data $\mathbf{x}_1, \dots, \mathbf{x}_n$. Attraction is quantified by $p_{j|i}$, defined as the conditional probability of moving from $\mathbf{x}_i$ to $\mathbf{x}_j$ in this graph. Next, we define a graph in the embedding space whose nodes are the embedded data $\mathbf{y}_1, \dots, \mathbf{y}_n$. For this embedding, we define an analogous conditional probability: $q_{j|i}$.

   To select an embedding, we minimize the following cost function:

   $$C = \sum_i KL(P_i \,||\, Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

   where

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

   and

   $$q_{j|i} = \frac{\exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2)}{\sum_{k \neq i} \exp(-\|\mathbf{y}_i - \mathbf{y}_k\|^2)}$$

   (a) Show that the gradient of C with respect to $\mathbf{y}_i$ is

   $$\frac{\partial C}{\partial \mathbf{y}_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(\mathbf{y}_i - \mathbf{y}_j)$$

2. Assume we have data $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$. Suppose that the indices $i = 1, \dots, n$ are split into two classes C_1 and C_2. For two-class linear discriminant analysis, we solve the following optimization problem:

   $$\max_{\mathbf{w} : \|\mathbf{w}\|=1} \frac{(\mu_1 - \mu_2)^2}{s_1^2 + s_2^2} = \max_{\mathbf{w} : \|\mathbf{w}\|=1} \frac{\mathbf{w}^\top S_b \mathbf{w}}{\mathbf{w}^\top S_w \mathbf{w}}$$

   $$S_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^\top$$

   and

   $$S_w = S_1 + S_2 = \sum_j \sum_{i \in C_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^\top$$

   where $\mathbf{m}_j$ represents the (high-dimensional) centroid of class j and μ_j represents the projected class mean for class j.

   (a) Show that S_w can be written as $X_c^\top X_c$ where where X_c is the $n \times p$ matrix where the ith row is given by the entries of $\mathbf{x}_i - \mathbf{m}_{j(i)}$ where the ith observation belongs to the $j(i)$th class.

(b) The optimization problem above cannot be solved directly if $n < p$. Explain why it cannot be solved and describe two possible ways to resolve this problem.

3. The emergence of manifold learning methods like UMAP and t-SNE has led to considerable debate about their usefulness for scientific discovery. Read the following papers by Chari and Pachter (2023) and Lause, Berens, and Kobak (2024+) and summarize the following (a brief paragraph should be enough for each). Note that there is no need to understand the single cell genomics applications completely; focus on what is being said about UMAP and t-SNE.

   (a) the main argument and conclusions of Chari and Pachter.

   (b) the main argument and conclusions of Lause, Berens, and Kobak.

   (c) any takeaways, ideas, and questions you have about manifold learning after reading these two papers.