**下发日期**: week 6, 2023.04.09

**提交日期**: 23:59 pm, week 11, 2023.05.14

**总分值**: 25

**提交邮箱**： data_visualization@yeah.net

**任务说明**：可视化至少两类具有高维表征的数据集，包括文本数据集（例如，Large Movie Review Dataset, http://ai.stanford.edu/~amaas/data/sentiment/）和图片数据集（例如，MNIST 或 CIFAR-10）。此外，鼓励物理、化学、生物等其他非计算机专业学生可视化自己学科相关高维数据集。使用不少于两类高维数据降维算法进行数据可视化，线性（例如，PCA）和非线性（例如，t-SNE）。同时，尝试调整模型超参，观察可视化结果变化。

**特别说明**：

（1）如果可视化两类要求数据集，则需要使用至少三种高维数据降维算法；

（2）如果可视化三类数据集，则需要至少使用两种高维数据降维算法。也就是，至少提交 6 幅可视化图片。

**提交内容**：推荐使用 python 等编程工具，不能使用 Excel。提交核心代码和可视化结果图或视频，并对不同方法获得的可视化结果进行对比和分析。

(English)

**Handout**: week 6, 2023.04.09

**Due**: 23:59 pm, week 11, 2023.05.14

**Total points**: 25

**Send to Email**: data_visualization@yeah.net

**Task**: Visualize at least two datasets with high-dimensional representations/embeddings, including text datasets (such as the Large Movie Review Dataset, http://ai.stanford.edu/~amaas/data/sentiment/) and image datasets (such as MNIST or CIFAR-10). Besides, students in disciplines such as Physics, Chemistry, and Biology are also encouraged to visualize their own high-dimensional datasets related to their research field. Use at least two high-dimensional data dimensionality reduction algorithms, both linear (such as PCA) and nonlinear (such as t-SNE), to perform visualization. Use different algorithm hyper-parameters to observe the result changes.

**Note:**

(1) If visualizing two types of datasets, you should use at least three high-dimensional data dimensionality reduction algorithms;

(2) If visualizing three types of datasets, two high-dimensional data dimensionality reduction algorithms should be used. In other words, at least 6 visualization images should be submitted.

**Submission**: Python or other programming platform is recommended. BUT, No Excel!

Please submit key source code, along with representative visualization results in the form of screenshot images or videos. Meanwhile, give experimental comparison and analysis for visualization results.