Data Wrangling & Visualization

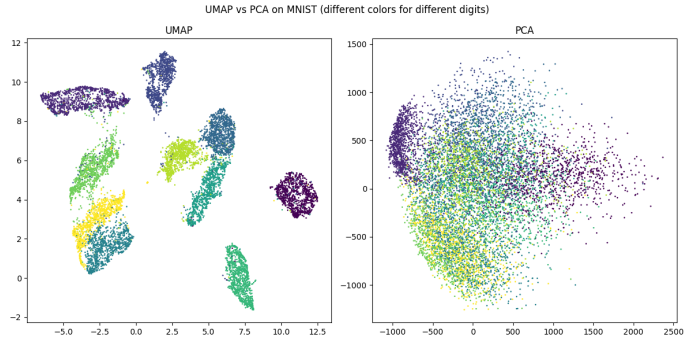
UMAP Visualization project

Nikita Zagainov, Dmitry Tetkin, Nikita Tsukanov

Innopolis April 2025

Background: Dimensionality Reduction

- A powerful tool to compress or simplify data
- The most advanced method nowadays - `UMAP` [1]



Different dimensionality reduction algorithms in action

[1] `https://arxiv.org/abs/1802.03426`

Project Goal

- To build a web application capable of providing a useful visualization of given high-dimensional dataset in `.csv` format
- To show the process of `UMAP` algorithm fitting to provided dataset
- To provide interactive tools to for exploration of low-dimensional representation of the dataset

How it works

Data preprocessing

We created a pipeline to preprocess the data and prepare it for visualization. This step is crucial as most of the datasets may contain non-numerical features, missing values, etc.

UMAP fitting

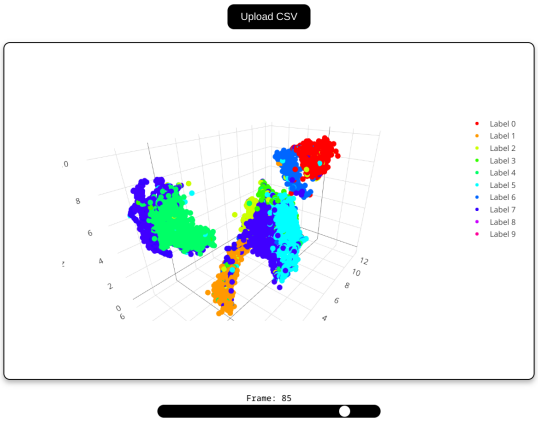
We apply modified version of original `UMAP` algorithm [2] to the preprocessed data to get low-dimensional representation of the dataset for 100 consecutive iterations.

Visualization app

We use `FastAPI` framework to deliver embeddings to the frontend application, where `Plotly` library is used to visualize the data.

[2] `https://github.com/lmcinnes/umap`

Demo



Demo of the web application

Useful links

Follow us on social media!

Project repository

Thank you for your attention!