

Dominika Jaśniewska

Aleksandra Warmbier

Report

On the beginning we load dataset to our program. Dataset is enormous so we take only a part of it. In other way after we run it, our programm returns memory errors. We try avoid this using batch processing smaller part of data but eventually we gave up. We use MDS method to reduce dimensionlity. MDS (multi dimensional scaling) is a distance-preserving manifold learning method. All manifold learning algorithms assume the dataset lies on a smooth, non linear manifold of low dimension and that a mapping $f: R^D \rightarrow R^d$ ($D \gg d$) can be found by preserving one or more properties of the higher dimension space. Distance preserving methods assume that a manifold can be defined by the pairwise distances of its points. In distance preserving methods, a low dimensional embedding is obtained from the higher dimension in such a way that pairwise distances between the points remain same. Some distance preserving methods preserve spatial distances (MDS) while some preserve graph distances. We present our result below.

