

# t-distributed stochastic neighbor embedding

Kristian Wichmann

May 23, 2017

## 1 Overview

*t-distributed stochastic neighbor embedding* - or t-SNE for short - is a dimensionality reduction algorithm. It takes a dataset in many dimensions and transforms it into a set with a lower dimensionality, usually two or three.

### 1.1 Terminology

Let  $x_1, x_2, \dots, x_N$  be a dataset in a high dimensional (euclidean) space. And let  $\sigma_1, \sigma_2, \dots, \sigma_N$  be a set of standard deviations of the same size.

## 2 First transformation

Let one of the data points  $x_i$  be given. Now consider the conditional probability of drawing  $x_j$  from a Gaussian distribution centered at  $x_i$  with standard deviation  $\sigma_i$ :