dimensionality reduction

Authors

Alexander Jung[*], Department of Computer Science, Aalto University, Espoo, Finland
[*]Corresponding author: alex.jung@aalto.fi

Abstract

Dimensionality reduction refers to methods that learn a map $h : \mathbb{R}^d \to \mathbb{R}^{d'}$ from a large set of raw features to a smaller set of $d' < d$ new features. Using fewer features lowers the effective dimension of a model, which reduces the danger of overfitting, and it reduces the computation needed for training. It also enables visualization: two learned features serve as the coordinates of a scatterplot of the data points. Principal component analysis (PCA) is the linear case, with the map chosen to minimize the reconstruction error, and an autoencoder learns a nonlinear map by the same criterion. A random projection reduces the dimension without any training, with distances approximately preserved.

Definition

Dimensionality reduction refers to methods that learn a transformation $h : \mathbb{R}^d \to \mathbb{R}^{d'}$ of a (typically large) set of raw features $x_1, \ldots, x_d$ into a smaller set of informative features $z_1, \ldots, z_{d'}$. Using a smaller set of features is beneficial in several ways:

- Statistical benefit: It typically reduces the risk of overfitting, as reducing the number of features often reduces the effective dimension of a model.
- Computational benefit: Using fewer features means less computation for the training of machine learning (ML) models. As a case in point, linear regression methods need to invert a matrix whose size is determined by the number of features.
- Visualization: Dimensionality reduction is also instrumental for data visualization. For example, a transformation can be learned that delivers two features z_1, z_2, which can then be used as the coordinates of a scatterplot. Fig. 1 depicts the scatterplot of handwritten digits that are placed using transformed features. Here, the data points are naturally represented by a large number of grayscale values (one value for each pixel).

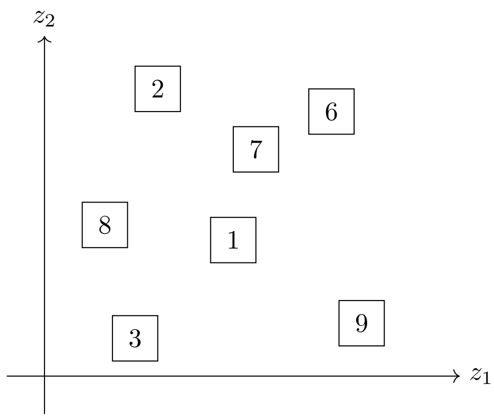

Figure 1: Example of dimensionality reduction: High-dimensional image data (e.g., high-resolution images of handwritten digits) embedded into 2-D using learned features (z_1, z_2) and visualized in a scatterplot.

Cross References

overfitting, autoencoder, random projection, PCA, JL lemma

References

1.