Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2021:12**
http://www.math.su.se

Non-Linear Dimensionality Reduction by an Information Theoretic Optimal Manifold Approach

Ruben Ridderström*

June 2021

Abstract

To visualize and understand high-dimensional data, performing some kind of dimensionality reduction is often required. Traditional methods such as principal-component analysis are not able to generalize well to non-linear data. Therefore non-linear methods are required. In this thesis we demonstrate such a method originally presented in the paper 'Optimal Manifold Representation of Data: An Information Theoretic Approach' by Chigirev and Bialek. This method uses information theoretic concepts to view dimensionality reduction as a compression problem to find the underlying manifold of the data. The method is nonparametric and has linear time complexity. There is one adjustable hyperparameter that allows us to look at the data structures at different spatial scales. In addition to demonstrating the method with intuitive test examples of clustering and manifold learning we also give a brief overview of fundamental information theoretic concepts and cover some basics of dimensionality estimation. We show that the method does manage to capture the underlying structure of the data both when performing clustering and in identifying underlying manifolds. And that it compares favorably to self-organizing maps which is a well-established method for performing dimensionality reduction.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: ruben.ridderstrom@gmail.com. Supervisor: Chun-Biu Li.