

StatMathAppli2017

Changepoint Detection in the Presence of Outliers

Paul Fearnhead¹ and Guillem Rigail²

¹Dept. of Mathematics and Statistics, Lancaster University

²LaMME & IPS2, INRA/Université d'Evry/CNRS

September 2017

Abstract

Many traditional methods for identifying changepoints can struggle in the presence of outliers, or when the noise is heavy-tailed. Often they will infer additional changepoints in order to fit the outliers. To overcome this problem, data often needs to be pre-processed to remove outliers, though this is difficult for applications where the data needs to be analysed online. We present an approach to changepoint detection that is robust to the presence of outliers. The idea is to adapt existing penalised cost approaches for detecting changes so that they use loss functions that are less sensitive to outliers. We argue that loss functions that are bounded, such as the classical biweight loss, are particularly suitable as we show that only bounded loss functions are robust to arbitrarily extreme outliers. We present an efficient dynamic programming algorithm that can find the optimal segmentation under our penalised cost criteria. Importantly, this algorithm can be used in settings where the data needs to be analysed online. We show that we can consistently estimate the number of changepoints, and accurately estimate their locations, using the biweight loss function. We demonstrate the usefulness of our approach for applications such as analysing well-log data and detecting copy number variation.

<https://arxiv.org/abs/1609.07363>