Sparse Maximin Aggregation of Neuronal Activity
Mogensen, Søren Wengel; Lund, Adam; Hansen, Niels Richard

Publication date:
2017

Document Version
Peer reviewed version

Citation for published version (APA):
Abstract—When analyzing large and inhomogeneous data sets it is of interest to obtain a robust estimate of an underlying signal. We consider a large data set describing neuronal activity in which systematic noise components are present. We propose the use of maximin aggregation and $L_1$ penalization to obtain a robust and sparse signal from this noisy data. An approximative computational method and an exact LARS-type method giving the entire solution path are presented.

I. INTRODUCTION

Let $X$ and $B$ be random vectors taking values in $\mathbb{R}^p$ and $\varepsilon$ be a zero-mean real random variable. Assume

$$Y = X^t B + \varepsilon.$$

We think of $X$ as a vector of predictor values and of $B$ as a vector of coefficients. Say now we have observations $Y_1, \ldots, Y_n$. If $B$ has a degenerate distribution such that with probability one $B = \beta$ for a $\beta \in \mathbb{R}^p$, we have a standard linear regression setting. If the distribution of $B$ is not degenerate we could still ask for a single $\beta \in \mathbb{R}^p$ to capture some feature of the data. For this purpose, define the maximin effects [1]

$$\arg \max_{\beta \in \mathbb{R}^p} \min_{\lambda \in \mathbb{R}} (2\beta^t \Sigma B - \beta^t \Sigma \beta),$$

where $\Sigma$ is the population Gram matrix of $X$ and $F$ is the support of the distribution of $B$. The maximin effects maximize minimal (over $F$) explained variance when compared to the constant prediction. We will use the term maximin aggregation to refer to the process of aggregating effects across $F$ to obtain the estimated maximin effects.

One can show that maximin aggregation enjoys a certain robustness property. Adding new vectors to $F$ will only bring the maximin effects closer to the origin which corresponds to the constant prediction. This feature makes them attractive to use on noisy and inhomogeneous data sets as false positive results are unlikely.

We will from now on only consider the case of known groups, meaning that a known partition of the set of observations is available such that the regression coefficient is constant within these groups. We enumerate these groups by the natural numbers $1, \ldots, G$ such that

$$Y_i = X_i^t B_{g(i)} + \varepsilon_i$$

for a labelling function $g : \{1, \ldots, n\} \rightarrow \{1, \ldots, G\}$.

To obtain a sparse result we add an $L_1$ penalization on the parameter vector,

$$\arg \min_{\beta \in \mathbb{R}^p} \max_g (\hat{V}_{\beta,b} + \lambda \|\beta\|_1)$$

(1)

where $\hat{V}_{\beta,b}$ is an empirical version of the explained variance and $\lambda$ is a non-negative tuning parameter.
Fig. 1. Snapshots of a single recording (raw data).

Fig. 2. Snapshots of the fitted maximin effects prediction.

Fig. 3. Measurements from a single recording device during several recordings. Top: 20 randomly selected tracks as they evolve over time. Middle: smoothed version of the 20 tracks above. Bottom: prediction by the maximin effects estimated from the full data set. Note the different scaling of the y-axes.

Fig. 4. Simple example of a solution path in $\mathbb{R}^2$. The three black line segments indicate the sets of points in which the loss is not differentiable. The square is the unpenalized maximin effects and the dashed line is the solution path. For large enough values of $\lambda$ the solution is the zero vector.

REFERENCES


