

PAPER 2:

Introduction:

1. Key contribution: An examination of the trade space for combinatorial and convex approaches to Robust Sparse PCA
2. Summary of candidates
 - a. Matt (exhaustive / greedy heuristic) (good with no noise)
 - b. Combinatorial (Deon, and Greedy Deon)
 - c. Convex
 - d. PCA is irrelevant
3. State key findings

Analysis

2.1 The synthetic data:

How was it made, what are the free parameters, type of tests which could be done on it

2.2 The methods

General approach: Steps of analysis (based on Randy, Melanie, brady discussions) – cf HealthRPCA_July2019_Brady_editted, slide 6 “Current approach”. How to find lambda

2.21 Matt: More details of approach, pseudo code, specific parameter identification

2.22 Combinatorial: Deon and Greedy Deon: more details, pseudo code.

- Variations: penalty for R and S, how to choose M, detecting bad M's how to choose which columns (greedy, guessing), quadratic or L1 objective, how to choose n' ,

2.23 Convex: more details, pseudo code.

- How to find lambda

2.3 Results

Map out results. Conduct experiments to map things out.

Metrix: RMS without anomalies, confusion matrices of anomalies, RMS vs number of prediction made

Experiments:

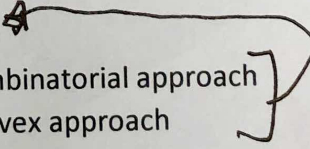
- Large and small problems
- Small problems which kills Deon
-

Health Paper

Primary references

- previous conference paper. Academy Health (2019)
- HealthRPCA_July2019_Brady_editted
- A 'matt' presentation
- Deon's writeup
- Melanie's thesis

1. Introduction

- Pull content from previous conference paper. Academy Health (2019) Les should find.
- Our contribution: We know and have enhanced methods which can solve a government problem.
- Government regulations impose burdens which should be minimized particularly if there is minimal information gained from added burden
- Sparse robust PCA solved many problems of real world context when there is a need to reduce government measure burden. This is accentuated for the case of poor quality data and in particularly sparse noisy data caused by human error. This is demonstrated in section 2 of this paper on synthetic and real dat
- We have enhanced existing work to create fast algorithms for solving this problem the details are in reference [1]
- 1-2 page summary of the methods
 - Modern methods can identify the reduced needs
 - Sparse PCA: (linear combination of a few columns)
 - No abnormals: "Matt" an exhaustive search or greedy algorithm.
 - Abnormals: 
 - Combinatorial approach
 - Convex approach
 - PCA *no abnormality, no noise*

2. Analysis

- Synthetic data. This is particularly useful to look at as we know the ground truth.
 - Describe general approach
 - What does the data look look
 - What are the high level steps to analyze it (1 page)
 - Results
 - PCA does terrible srt to RPCA and Gurobi
 - High level difference between RPCA and Gurobib
 - Real data
 - What is the data
 - What does PCA/ RPCA/Gurobi find
 - Do Paffenroth insertions: add anomalies and show that we found it.