CMSC426HW1 Report

Yizhan, Ao
josephao@umd.edu
UID:116022064

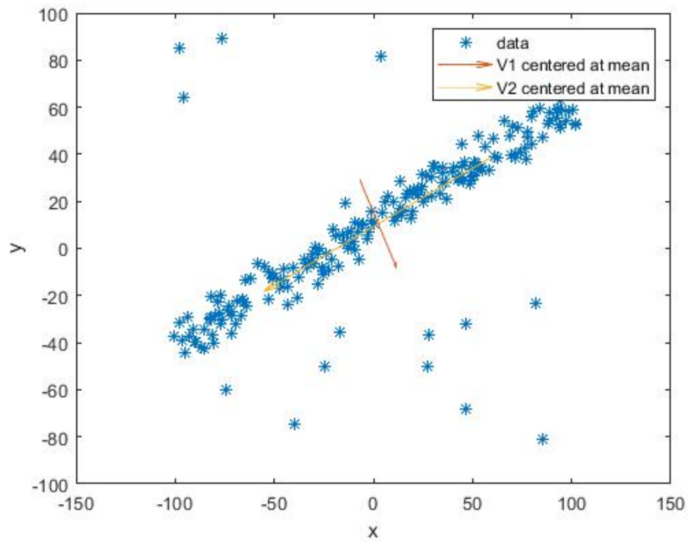
September 20nd 2021

1 Introduction

- In this report you will be presented with matlab code visualization of geometric representation of Eigenvalues and covariance matrices
- The outlier rejection of each datasets with arguing of the optimal choices
- Each rejection technique pros and cons

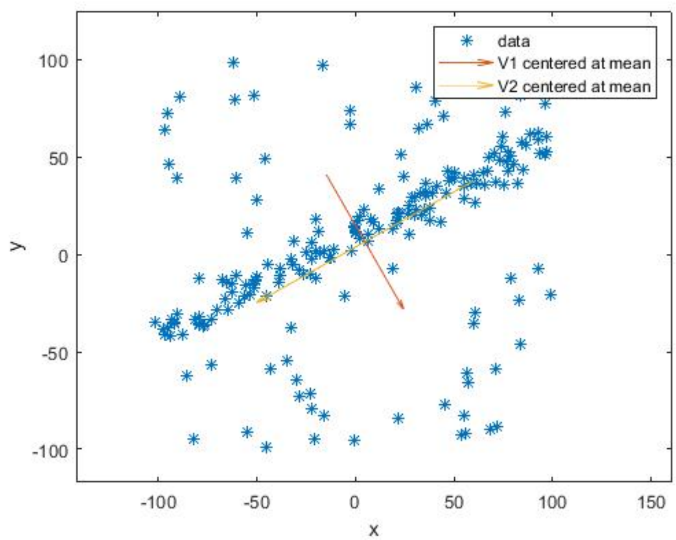
2 Eigenvalue and Eigenvectors

- Eigenvectors are vectors that keep their direction after being transformed linearly by any matrix (A).
- Mathematically speaking, $A * v = \lambda * v$ where λ lies as a scalar. Also known as Eigenvector.
- During A^T, eigenvector can find the rotation of a dataset therefore not changing its rotation after the linear transformation



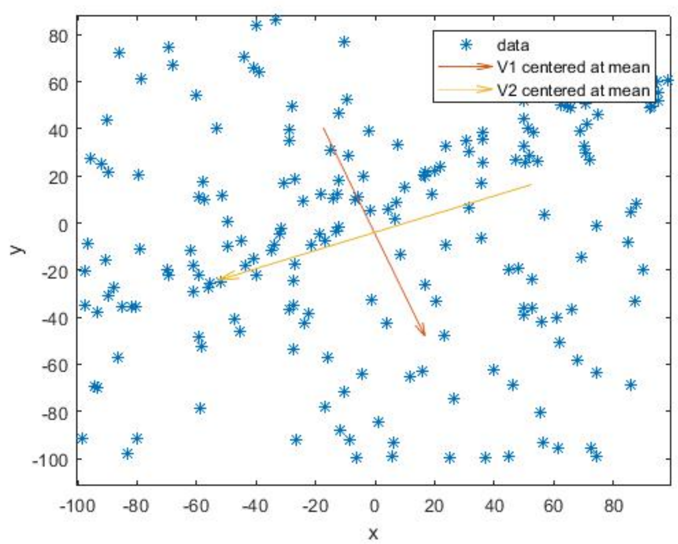
1. From the dataset 1 we can see the spread and the verticla speed of the data couldnt explain the correlation. Therfore we have to use the covariance of the matrix to explain the magnitude and the direction. The Eigenvectors from the set are having a linear relationship while getting a closer look on the pattern, One arrow is pointing down and one is pointing left down, According to the description of the Problem we have a conclusion that the data is having a low noise.

2. The dataset 2 has a moderate noise given a little bit higher coordinates and the linear regression couldn't have an accurate fit. By tuning the noise of the function parameter λ we can have a better fit of the model which will need to choose the threshold. RANSAC, also provides a good

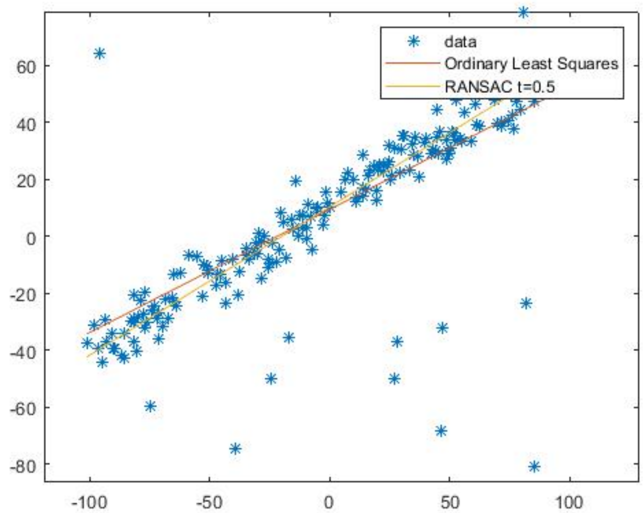


fit.

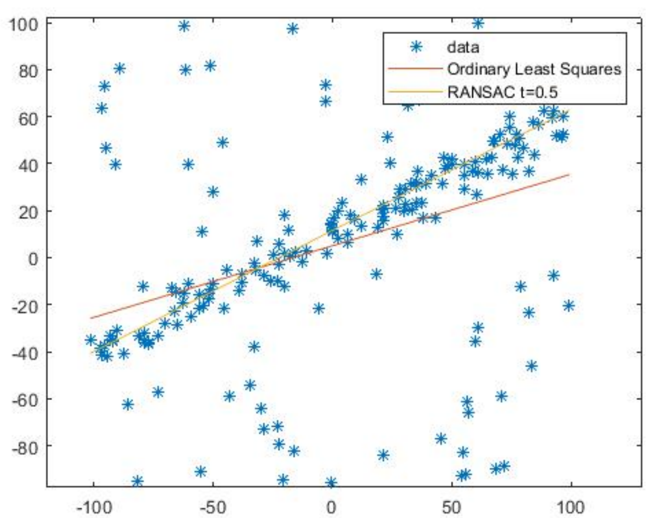
3. For dataset3 which has a much higher noise and liunear regression is more spread out given coordinates. RANSAC, on the other hand, provides remarkably accurate results when it comes to selecting the appropriate parameters such as inlier threshold, inlier ration, and so on.



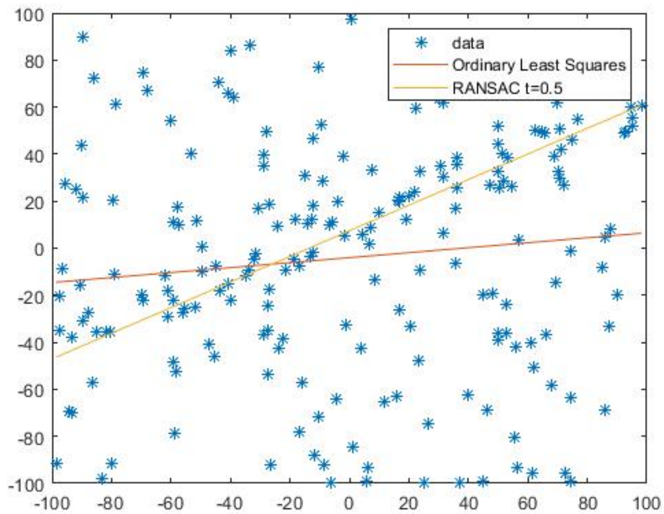
3 Optimal Choices



- The RANSAC approach is my preferred method since it provides the most visually beautiful fit. However, in terms of algorithm efficiency, I believe the linear method provides a better fit.
- From Dataset 2 we have more noise compared to the last dataset. I would say that although Least Squares lines have the computational efficiency advantage, they do not combat the noise in data 2 well enough



- For Dataset 3, there are much more noise and more biases. I would say the RANSAC method outperform the Ordinary Least Square in this data set. The Ordinary Least Squares is strongly affected by the noise in the right bottom of the scatter plot. We need to zoom out from the picture to see the whole data points on the picture. Please forgive me for the bad zooming.



4 Limitation of each algorithm:

- linear Regression works well when the noise is low and the level of boise is low
- Regularization has more biance exchanging with variance. Given a dataset which has a low bias but high variance it couldnt do good representation since every possible outcome has a good fitr using least square.
- The RANSAC algorithm is overly reliant on the available computing complexity. Such that, when it comes to determining the problem-specific distance threshold, RANSAC requires more human interaction. Even by human observations, there is no absolute way that which threshold is the most optimal time saver.

5 References

- https://cmsc426.github.io/math-tutorial/
- https://www.visiondummy.com/2014/04/geometric-interpretation-covariance-matrix/

    - https://en.wikipedia.org/wiki/Random$_{s}ample_{c}onsensus : \;: text = Randomhttp :$
      $//www.cse.yorku.ca/\ kosta/CompVis_{N}otes/ransac.pdf$

- https://scikit-image.org/docs/dev/auto$_{e}xamples/transform/plot_{r}ansac.htmlhttps :$
  $//medium.com/@angel.manzur/got-outliers-ransac-them-f12b6b5f606e$