Bruxelles, June 2019

Least squares regression
a short story on overdetermined linear systems and Moore-Penrose pseudoinverse

Kamila Zdybał

Université libre de Bruxelles, kamila.zdybal@ulb.ac.be
camillejr.github.io/science-docs, kamila.zdybal@gmail.com

Preface

Imagine a linear system of equations with more number equations than the number of unknowns.

Goal of this paper: explain why Moore-Penrose inverse give a least squares regression.
This document is still in preparation. Please feel free to contact me with any suggestions, corrections or comments.

Keywords

overdetermined linear systems, partial least squares regression, Moore-Penrose inverse

Contents

1 References on Least Squares

2 Least Squares - the story

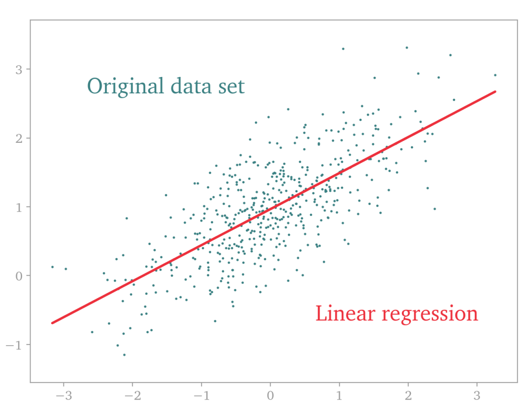

Figure 1: Example of a linear regression.

3 Partial Least Squares (PLS) regression

The most general linear equation is of the form:

$$\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \tag{1}$$

At this point, I ought to give a small justification. In many literature references such as [1], the above equation can be encountered written as $A\vec{x} = \vec{b}$ (with unknown $\vec{x}$) or $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ (with unknown $\boldsymbol{x}$). I decided to write it in a different notation, with the unknown vector $\boldsymbol{w}$, since I wanted to highlight the link of algebra with geometry.

To avoid naming confusion, in eq.(1), $\boldsymbol{X}$ and $\boldsymbol{y}$ contain the actual x and y coordinates of the data points on the x and y axis respectively. Thus, $\boldsymbol{X}$ and $\boldsymbol{y}$ are known; the vector of unknown coefficients is $\boldsymbol{w}$.

This actually follows the notation presented in [2] in Chapter 3 on *Linear Models for Regression*.

2D data example

Have a data set in pairs: (X, Y). X is a vector of N points and Y is a vector of corresponding N points. Together they make a cloud of points on a 2D-plane.

We now say that: X A = Y.

We are interested in finding the coefficients: y = C x + D.

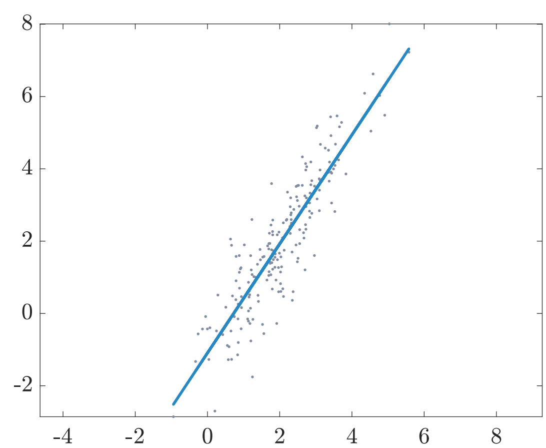

Figure 2: Linear basis function LS regression.

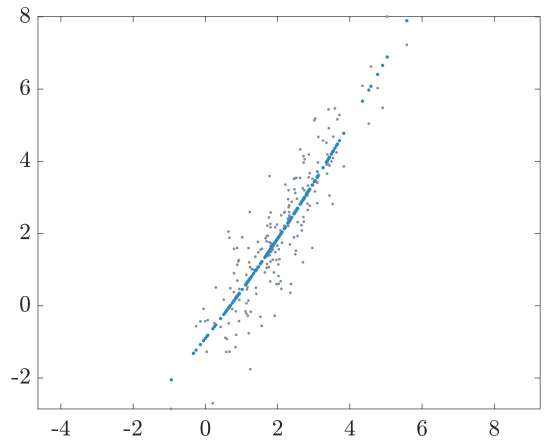

Figure 3: Non-linear (quadratic) basis function LS regression.

References

[1] Gilbert Strang, *Introduction to Linear Algebra*, Fifth Edition, 2016

[2] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, 2006

[3] Nathan Kutz, *Data Driven Discovery of Dynamical Systems and PDEs*, an online lecture

[4] T. Hastie, R. Tibshirani, J. Friedman, *Elements of Statistical Learning, Data Mining, Inference, and Prediction*, Second Edition, 2008