Lab6: Ordinary Least Squares Fitting

STUDENT NAME

February 21, 2016

1 Objective

The objective of this lab is to introduce the concept of Ordinary Least Squares fitting, and to learn how to program this algorithm in MatLab. En passé the student will learn general methods of matrix manipulation using MatLab.

2 Introduction

One of the most common forms of data analysis is to discover underlying themes in data. Often we use 'best fit' models where we assume that the mechanism that connects an input to an output can be described by a simple mathematical relationship (such as the polynomial) and we hope that the remainder (residuals) are random noise. A good example is a flow rate versus a pressure, where we make the assumption that Bernoulli's law holds, and therefore there should be a quadratic relationship between pressure and flowrate. The aim is to explain most of the data using the fewest number of terms (this is called a parsimonious model). The general polynomial model we are going to fit on the data can be written as follows:

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + e \tag{1}$$

where e is an error term. In matrix form:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & \vdots & \vdots \\ 1 & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \tag{2}$$

or

$$\bar{y} = A\bar{\theta} + \bar{e} \tag{3}$$

To give us the best fitting polynomial, we minimize the square of the error which gives us an Ordinary Least Squares (OLS) estimator and obtain the optimal parameter vector as follows:

$$\min(e^T e) = min \left[\left(\bar{y} - A\bar{\theta} \right)^T \left(\bar{y} - A\bar{\theta} \right) \right] \tag{4}$$

The result is:

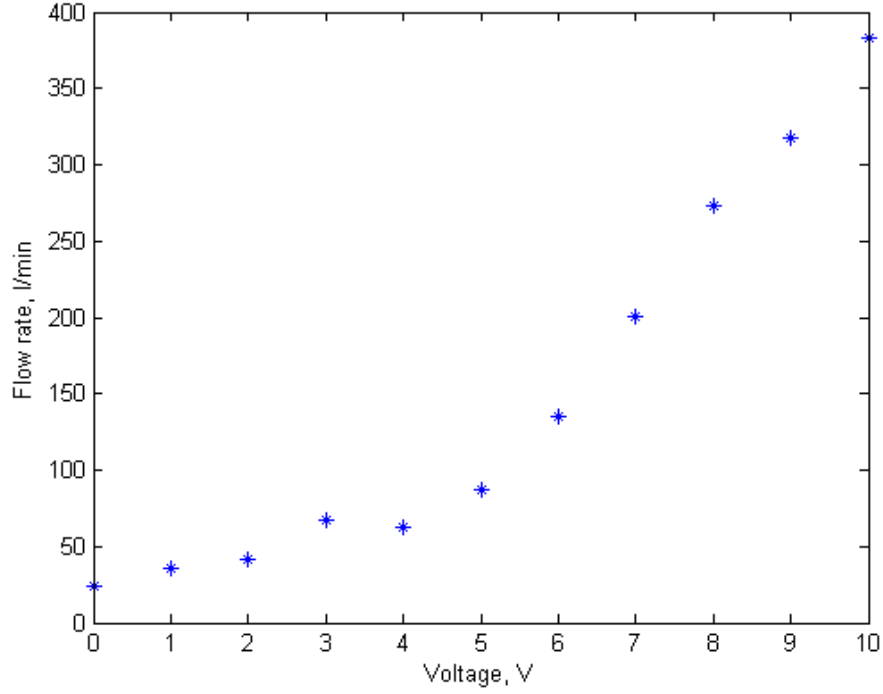$$\hat{\bar{\theta}} = \left(A^T A \right)^{-1} A^T y \tag{5}$$

Figure 1: These data represent the relationship between an electric input (x-axis, Volt) and a flow rate output (y-axis, l/min).

The coefficient of determination R^2 indicates how good the fit is. For this we need two factors, 1) the sum of squared residuals (SSR), which is fact the sum of the squared differences between the data and the model (this is our $S = e^2 = \sum_i e_i^2 = \sum_i (y_i - f(x,\theta))^2$) and 2) the total sum of squares of the dataset as follows: $SST = \sum_i (y_i - \bar{y})^2$, where $\bar{y}$ is the mean of y. The coefficient of determination is now:

$$R^2 = 1 - \frac{SSR}{SST} \tag{6}$$

This formula shows you how you can compute the coefficients based on the regressor matrix A and the data vector y, which contains the measurements.

3   Procedures

If you are not familiar (or somewhat rusty) in MatLab, do the Questions section first. Given is data vector x,y as follows (and shown in Figure 1):

2

$$x = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{pmatrix} \quad y = \begin{pmatrix} 24 \\ 36 \\ 42 \\ 67 \\ 62 \\ 87 \\ 135 \\ 201 \\ 273 \\ 318 \\ 383 \end{pmatrix} \tag{7}$$

1. In MatLab, build the regressor matrix

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & \vdots & \vdots \\ 1 & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \tag{8}$$

2. Compute the coefficient vector $\hat{\bar{\theta}} = \left(A^T A\right)^{-1} A^T y$ in a single MatLab statement.

3. Plot the original data.

4. Plot the model as $\bar{y} = A\bar{\theta}$ in the same figure. Type "help hold" for this.

5. Plot the error in a separate figure.

6. Compute the square of the error $e^T e$.

7. Build a generic routine where the user can select which terms are going to be in the model and which are not. Think about how to set up your regressor matrix related to the Order Vector Ovec as described in the template Figure 2.

4 Questions

Q1: What is the dimension of a product of an n*m matrix times an m*1 vector?
A1:

Q2: What is the dimension of a product of a 1*m vector times an m*1 vector? What is the name of a matrix with this dimension?
A2:

Q3: What is the dimension of a product of an m*1 vector times a 1*m vector
A3:

Q4: Sometimes instead of a vector product, we simply want to multiply all the entries of the vector with each other, resulting in a vector of the same dimension as the original vectors. How does one

```matlab
function [theta, msq] = fitols(x,y,Ovec)
% Fit polynomial function on data in OLS sense
% Author   :
% Date     :
% Revision :
%
% Syntax    : [theta,msq] = fitols(x,y,Ovec)
%
% theta     :  Parameter vector
% msq       :  Mean square error
% x         :  Independent variable
% y         :  Dependent variable
%
%      Ovec indicates terms [1 x x^2 ..]*Ovec'
%      Example Ovec = [1 0 1] gives [1 x^2] and not x

% If vectors x,y are horizontal, transpose them to make them vertical


% Make sure the x and y vector have the same length. If not alert the user
% with an error dialog box (type help errordlg )


% Build the matrix of regressors. Check each entry of Ovec, and if it is a
% 1, add another column to the regression matrix A.
A = [];


% Compute the parameter vector theta using the OLS formula


% Compute the error vector


% Compute the Coefficient of Determination. For this you need the Sum of
% Squared Residuals (SSR), and the Total Sum of Squares (SST).


% Plot y (pressure in MPa) versus x (Voltage, V). Add labels and title.
% Your output should look as shown in the handout.
|
```

Figure 2: Template for the fitols function, download this file in plain text format here, cut and past into MatLab's editor. You can also download the data here. If all terms are in the model (Ovec = [1 1 1]) the output should look like Figure 3
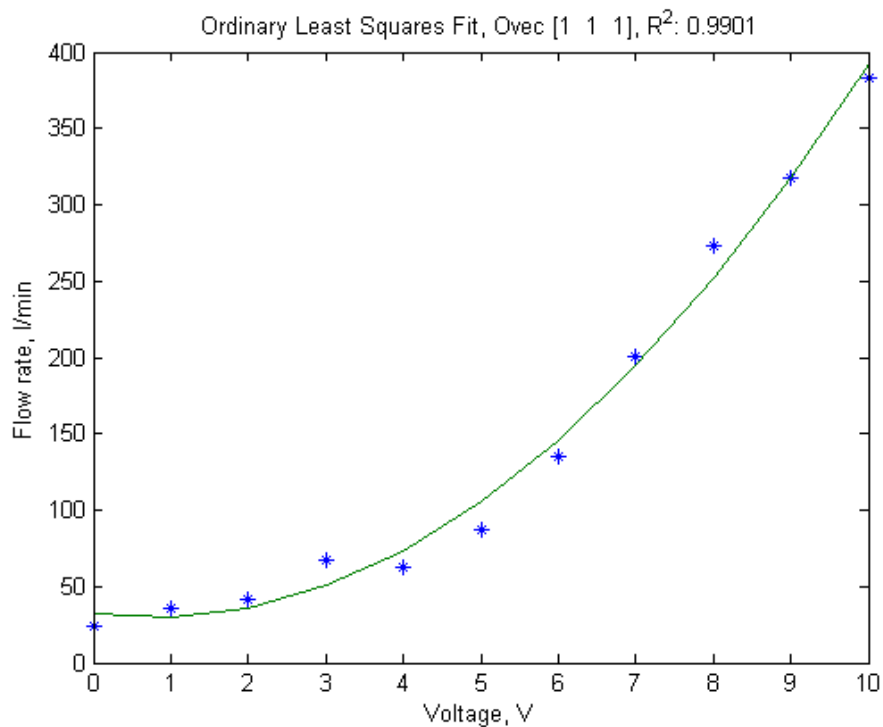
Figure 3: These data represent the relationship between an electric input (x-axis, Volt) and a flow rate output (y-axis, l/min). The fit parameters are shown in the title.

do this in MatLab?
A4:

Q5: Given are the following matrices and vector.

$$a = \begin{pmatrix} 4 & 9 & 2 \\ 3 & 5 & 7 \\ 8 & 1 & 6 \end{pmatrix} \quad b = \begin{pmatrix} 7 & 9 & 3 \\ 3 & 6 & 2 \\ 5 & 1 & 4 \end{pmatrix} \quad c = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} \tag{9}$$

Compute:

$$a + b = \tag{10}$$
$$a * c = \tag{11}$$
$$a^T = \tag{12}$$
$$a^{-1} = \tag{13}$$
$$c^T * c = \tag{14}$$
$$c * c^T = \tag{15}$$
$$c. * c = \tag{16}$$
$$(a^T a)^{-1} a^T c = \tag{17}$$
$$\tag{18}$$

A5: