

# Polynomial Regression (Handwriting Assignment)

Name: 강민성

Student ID: 20234424

Instructor: Professor Kyungjae Lee

October 9, 2023

---

## Introduction

In the mid-term project, we will look at a polynomial regression algorithm which can be used to fit non-linear data by using a polynomial function. The polynomial Regression is a form of regression analysis in which the relationship between the independent variable  $x$  and the dependent variable  $y$  is modeled as an  $n$ th degree polynomial in  $x$ .

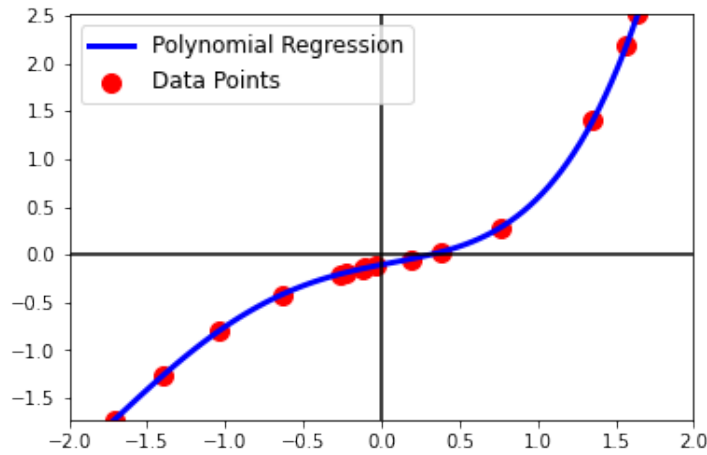


Figure 1: Example of Polynomial Regression

First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression can be defined as *finding a function that best explain data which consists of input and output pairs.* Let assume that we have 100 data points,

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_{98}, y_{98}), (x_{99}, y_{99}), (x_{100}, y_{100}).$$

The goal of regression is to find a function  $\hat{f}$  such that

$$\hat{f}(x_1) = y_1, \hat{f}(x_2) = y_2, \hat{f}(x_3) = y_3, \dots, \hat{f}(x_{99}) = y_{99}, \hat{f}(x_{100}) = y_{100}.$$

This is the simplest definition of the regression problem. Note that many details about regression analysis are omitted here, but, you will learn more rigorous definition in other courses such as



Figure 2: Examples of polynomial functions

machine learning or statistics. Then, the polynomial regression is the regression framework that employs the polynomial function to fit the data.

So, what is the polynomial function? I guess you may remember, from high school, the following functions:

$$\text{Degree of 0 : } f(x) = w_0$$

$$\text{Degree of 1 : } f(x) = w_1 \cdot x + w_0$$

$$\text{Degree of 2 : } f(x) = w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\text{Degree of 3 : } f(x) = w_3 \cdot x^3 + w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\vdots$$

$$\text{Degree of } d : f(x) = \sum_{i=0}^d w_i \cdot x^i,$$

where  $w_0, w_1, \dots, w_d$  are a coefficient of polynomial and  $d$  is called a degree of a polynomial. So, we can determine a polynomial function  $f(x)$  by deciding its degree  $d$  and corresponding coefficients  $\{w_0, w_1, \dots, w_d\}$ . Figure 2 illustrates some examples of polynomial functions.

Then, the polynomial regression is a regression problem to find the best polynomial function to fit the given data points. Especially, the polynomial function is determined by coefficients (let just assume that  $d$  is fixed). We can restate the polynomial regression as *finding coefficients of polynomials such that, for all data point,  $(x_i, y_i)$ ,  $y_i = \hat{f}(x_i)$  holds* (if we have noise free data). Figure 1 shows the example of polynomial regression. In the following problems, you have to study how to compute the coefficients of the polynomial to fit the data points.

## Problems

### 1. (80 pt. in total)

Assume that we have  $n$  data points,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . Let the degree of polynomial be  $d$ . Then, we want to find  $w_0, w_1, w_2, \dots, w_d$  of the polynomial such that

$$\begin{aligned}\hat{f}(x_1) &= w_0 + w_1x_1 + w_2x_1^2 + \dots + w_dx_1^d = y_1, \\ \hat{f}(x_2) &= w_0 + w_1x_2 + w_2x_2^2 + \dots + w_dx_2^d = y_2, \\ \hat{f}(x_3) &= w_0 + w_1x_3 + w_2x_3^2 + \dots + w_dx_3^d = y_3, \\ \hat{f}(x_4) &= w_0 + w_1x_4 + w_2x_4^2 + \dots + w_dx_4^d = y_4, \\ \hat{f}(x_5) &= w_0 + w_1x_5 + w_2x_5^2 + \dots + w_dx_5^d = y_5, \\ &\vdots \\ \hat{f}(x_n) &= w_0 + w_1x_n + w_2x_n^2 + \dots + w_dx_n^d = y_n.\end{aligned}$$

Now, we reformulate the equations into the vector and matrix form. First, let  $\mathbf{w} = [w_0, w_1, \dots, w_d]^T$  and  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ . Then, the above equations can be rewritten as

$$\hat{f}(x_1) = [1, x_1, x_1^2, x_1^3, \dots, x_1^d] \cdot \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_d \end{bmatrix} = [1, x_1, x_1^2, x_1^3, \dots, x_1^d] \mathbf{w} = y_1$$

Similarly, we have,

$$\begin{aligned}[1, x_2, x_2^2, x_2^3, \dots, x_2^d] \mathbf{w} &= y_2, \\ [1, x_3, x_3^2, x_3^3, \dots, x_3^d] \mathbf{w} &= y_3, \\ [1, x_4, x_4^2, x_4^3, \dots, x_4^d] \mathbf{w} &= y_4, \\ [1, x_5, x_5^2, x_5^3, \dots, x_5^d] \mathbf{w} &= y_5, \\ &\vdots \\ [1, x_n, x_n^2, x_n^3, \dots, x_n^d] \mathbf{w} &= y_n.\end{aligned}$$

Then, all equations can be written as the form of linear equation,

$$A\mathbf{w} = \mathbf{y},$$

where  $A$  is the stack of  $[1, x_i, x_i^2, x_i^3, \dots, x_i^d]$  for  $i = 1, \dots, n$ . Under this setting, answer the following questions.

1-(a) What is the size of vector  $w$  and  $y$ ? (10pt)

$$w = d+1$$

$$y = n$$

1-(b) What is the size of matrix  $A$ ? Write  $A$ . (10pt)

$$Aw = y$$

$$w = (d+1) \times 1$$

$$y = n \times 1$$

so size of matrix  $A$   
is  $n \times (d+1)$

1-(c) Let  $d+1 = n$ , then,  $A$  becomes a square matrix. Compute the determinant of  $A$ . (40pt in total, Derivation: 30pt, Answer: 10pt)

Let determinant of  $A$  is  $V_n$ .

$$V_n = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix}$$

By multiple of Row added to Row of Determinant.

$$V_n = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & \dots & x_2^{n-1} - x_1^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n - x_1 & x_n^2 - x_1^2 & \dots & x_n^{n-1} - x_1^{n-1} \end{vmatrix}$$

$x_1$  times column  $n-1$  from column  $n$ ,

$x_1$  times column  $n-2$  from column  $n-1$ .

$$a_{jj} = (x_2^{j-1} - x_1^{j-1}) - (x_1 x_2^{j-2} - x_1^j) = (x_2 - x_1) x_2^{j-2}$$

$$V_n = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & x_2 - x_1 & (x_2 - x_1) x_2 & \dots & (x_2 - x_1) x_2^{n-2} \\ 0 & x_3 - x_1 & (x_3 - x_1) x_3 & \dots & (x_3 - x_1) x_3^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n - x_1 & (x_n - x_1) x_n & \dots & (x_n - x_1) x_n^{n-2} \end{vmatrix}$$

$$V_n = \prod_{k=2}^n (x_k - x_1) \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & x_2 & \dots & x_2^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & x_n & \dots & x_n^{n-2} \end{vmatrix} = \prod_{k=2}^n (x_k - x_1) \begin{vmatrix} x_2 & \dots & x_2^{n-2} \\ x_3 & \dots & x_3^{n-2} \\ \vdots & \vdots & \vdots \\ x_n & \dots & x_n^{n-2} \end{vmatrix}$$

$$V_n = \prod_{k=2}^n (x_k - x_1) V_{n-1}$$

$V_2$ , by the time we get to it (it will concern elements  $x_{n-1}$  and  $x_n$ )

$$V_2 = \begin{vmatrix} 1 & x_{n-1} \\ 1 & x_n \end{vmatrix} = x_n - x_{n-1}$$

$$\text{So, } |A| = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$

1-(d) What is the condition that makes the determinant of  $A$  non-zero? (10pt)

$x_i (1 \leq i \leq n)$  are all different.

1-(e) Assume that the determinant of  $A$  is non-zero, then, what is the solution of linear equation,  $Aw = y$ , with respect to  $w$ ? (10pt)

If  $\det A \neq 0$ , there is matrix  $B$  that  $BA = I$   
and  $AB = I$  which is  $A^{-1}$ .

$$\text{so, } Aw = y$$

$$A^{-1}Aw = A^{-1}y$$

$$w = A^{-1}y$$

## 2. (20pt)

Suppose that  $n > d + 1$ . Then, we cannot compute the inverse of  $A$  since  $A$  is not a square matrix. In this case, how can we solve the linear equation  $Aw = y$ ?

Let  $A^+$  be a pseudo inverse of  $A$  and  $A$  is  $n \times (d+1)$  Vandermonde matrix ( $n > d+1$ )

By the definition of pseudo inverse, following properties hold.

- $AA^+A = A$
- $A^+AA^+ = A^+$
- $A^+ = (A^TA)^{-1}A^T$

So, if column vectors of  $A$  are linearly independent, we can compute  $w = A^+y$

$$\begin{pmatrix} Aw = y \\ A^+Aw = A^+y \\ (A^TA)^{-1}A^TAw = A^+y \\ w = A^+y \end{pmatrix}$$

$A$  is  $n \times (d+1)$  Vandermonde matrix, so

By the SVD.  $A = U\Sigma V^T$  ( $U$  is  $n \times n$  orthogonal matrix,  
 $V$  is  $(d+1) \times (d+1)$  orthogonal matrix  
 $\Sigma$  is  $n \times (d+1)$  diagonal matrix)

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{d+1} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_n \quad (\sigma_i \text{ are singular value})$$

$$A^+ = V\Sigma^+U^T, \quad \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_{d+1} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\text{and } A = \begin{bmatrix} x_1^0 & x_1^1 & x_1^2 & \dots & x_1^n \\ x_2^0 & x_2^1 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^0 & x_n^1 & x_n^2 & \dots & x_n^n \end{bmatrix}$$

So let assume that  $C_0v_0 + C_1v_1 + \dots + C_nv_n = \vec{0}$ , where  $v_j = (x_0^j, x_1^j, \dots, x_n^j)$

is the  $j$ -th column written as a vector and  $C_0, \dots, C_n \in \mathbb{R}$

Then we can get the  $k$ -th coordinate

$$C_0 + C_1x_k + C_2x_k^2 + \dots + C_nx_k^n = 0$$

which means that  $x_k$  is a root of the polynomial  $p(x) = C_0 + C_1x + C_2x^2 + \dots + C_nx^n$

Now is the polynomial  $p(x)$  of degree at most  $n$  has  $(n+1)$  different roots  $x_0, x_1, \dots, x_n$ ,

it must be the zero polynomial and we get that  $C_0 = C_1 = \dots = C_n = 0$ .

So, the vectors  $v_0, v_1, v_2, \dots, v_n$  are linearly independent.

Therefore,  $w = A^+y$