

Polynomial Regression (Handwriting Assignment)

Name: Benarache Hae

Student ID: 50231642

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 .

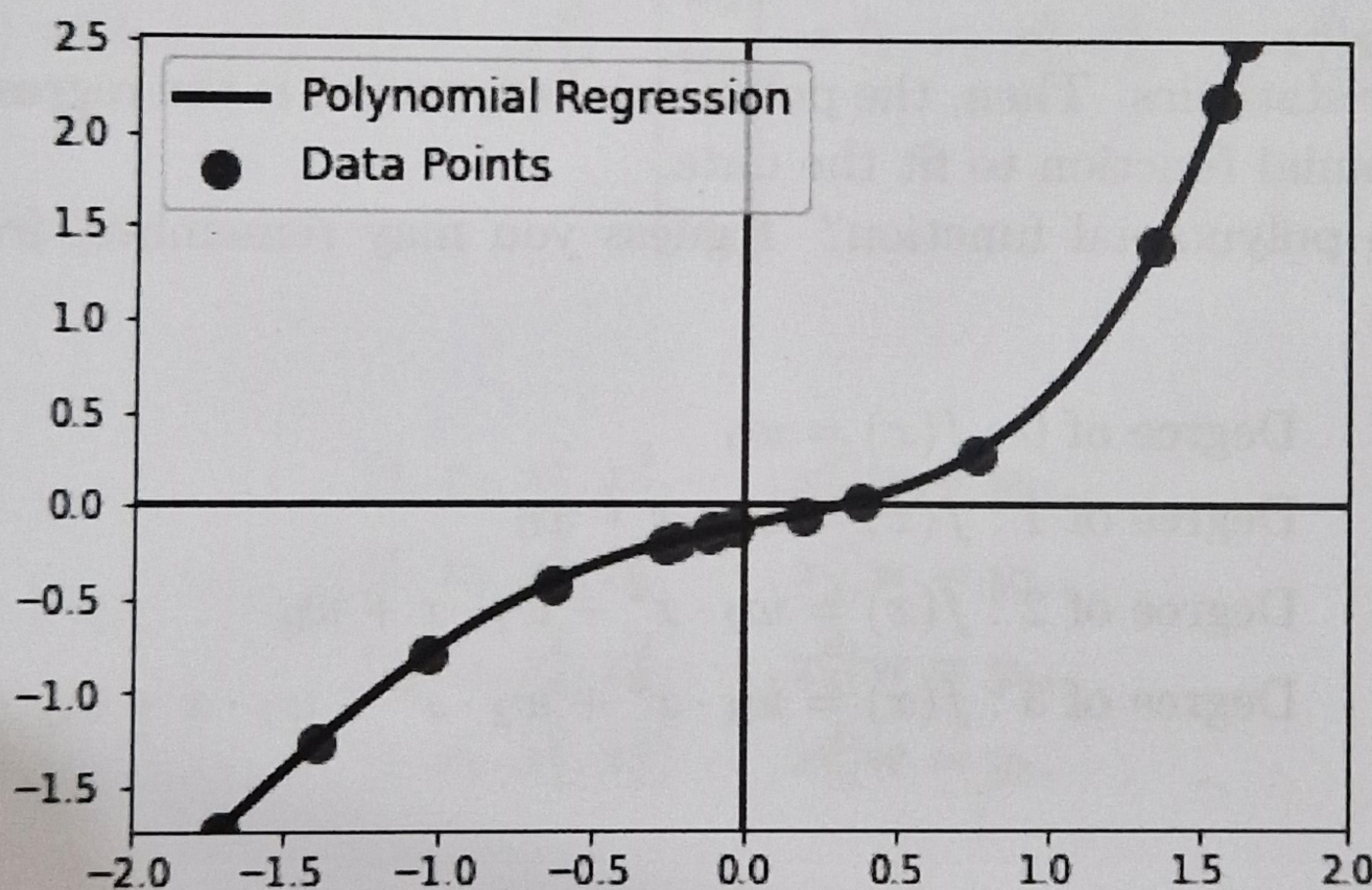


Figure 1: Example of Polynomial Regression

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

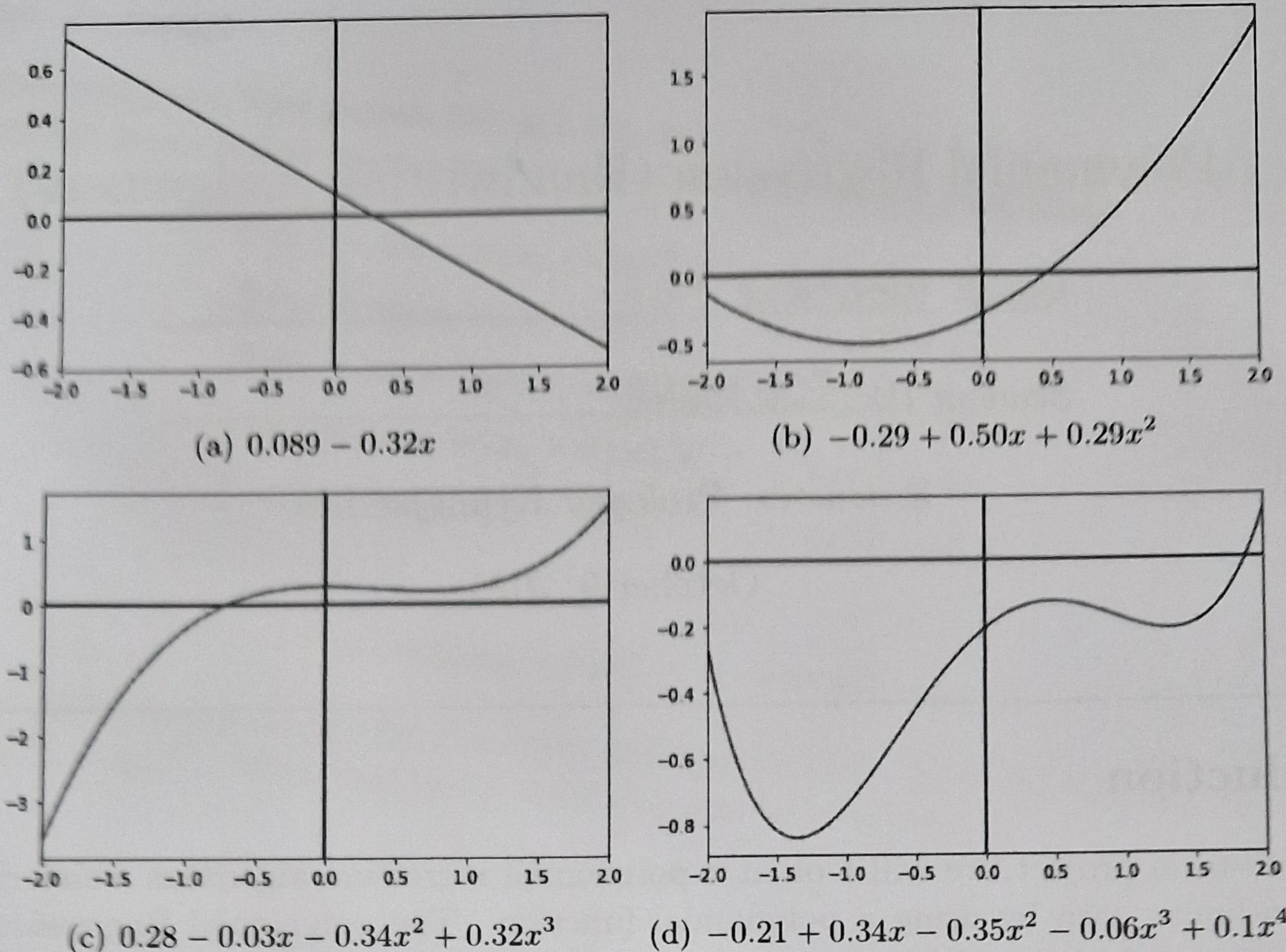


Figure 2: Examples of polynomial functions

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:

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

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

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

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

⋮

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

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

$$\begin{aligned}\hat{f}(x_1) &= w_0 + w_1 x_1 + w_2 x_1^2 + \dots + w_d x_1^d = y_1, \\ \hat{f}(x_2) &= w_0 + w_1 x_2 + w_2 x_2^2 + \dots + w_d x_2^d = y_2, \\ \hat{f}(x_3) &= w_0 + w_1 x_3 + w_2 x_3^2 + \dots + w_d x_3^d = y_3, \\ \hat{f}(x_4) &= w_0 + w_1 x_4 + w_2 x_4^2 + \dots + w_d x_4^d = y_4, \\ \hat{f}(x_5) &= w_0 + w_1 x_5 + w_2 x_5^2 + \dots + w_d x_5^d = y_5, \\ &\vdots \\ \hat{f}(x_n) &= w_0 + w_1 x_n + w_2 x_n^2 + \dots + w_d x_n^d = y_n.\end{aligned}$$

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 is composed of $d+1$ rows 1 column

$\rightarrow w$ is a $d+1$ size vector

$$w = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix}$$

y is composed of n rows and 1 column

$\rightarrow y$ is a n size vector

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}$$

\Rightarrow The size of vector w is $d+1$ and y is n

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

A is a stack of $[1, x_i, x_i^2, \dots, x_i^d]$ for $i = 1, \dots, n$ $Aw = y$

So

$$A = \left(\begin{array}{cccc} 1 & x_1 & x_1^2 & \dots & x_1^d \\ 1 & x_2 & x_2^2 & \dots & x_2^d \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^d \end{array} \right) \quad \left. \right\}^{n \text{ rows}}$$

The 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 $d+1 = n \Rightarrow A$ is a square matrix

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{pmatrix}$$

We can see that A is a matrix of Vandermonde because:

$\rightarrow A$ is a square matrix with the same number of rows and columns

\rightarrow the entries in each row are the powers of x_i

\Rightarrow So A is a matrix of Vandermonde

We can use the properties of this matrix to compute the determinant of A

The Vandermonde matrix V has the following property to compute his determinant:

"The determinant of a Vandermonde matrix can be expressed as the product of the differences between the x_i values"

So we have: $\det(A) = \prod_{0 \leq i < j \leq n} (x_j - x_i)$

$$\det(A) = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3) \dots (x_n - x_1)(x_n - x_2) \dots (x_n - x_{n-1})$$

Proof:

$$\text{Let } B = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}$$

$$\det(B) = \det \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix} = \det \begin{pmatrix} 1 & x_1 & x_1^2 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 \end{pmatrix} = (x_2 - x_1)(x_3 - x_1) \det \begin{pmatrix} 1 & x_1 & x_1^2 \\ 0 & 1 & x_1 + x_2 \\ 0 & 1 & x_1 + x_3 \end{pmatrix}$$

$$\det(B) = (x_2 - x_1)(x_3 - x_1) \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & x_2 \\ 0 & 1 & x_3 \end{pmatrix} = (x_2 - x_1)(x_3 - x_2) \det \begin{pmatrix} 1 & x_2 \\ 1 & x_3 \end{pmatrix} = \prod_{1 \leq i < j \leq 3} (x_j - x_i)$$

We can apply the same method for the matrix A

$$\det(A) = \det \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{pmatrix} = \det \begin{pmatrix} 1 & x_1 & x_1^2 & & & & & & & \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & & & & & & & \\ \vdots & \vdots & \vdots & \ddots & & & & & & \\ 0 & x_n - x_1 & x_n^2 - x_1^2 & \dots & x_n^{n-1} - x_1^{n-1} & & & & & \end{pmatrix}$$

$$\det(A) = \prod_{2 \leq j \leq n} (x_j - x_1) \det \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & x_2^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & x_{n-1}^{n-2} \end{pmatrix}$$

$$\Rightarrow \boxed{\det(A) = \prod_{0 \leq i < j \leq n} (x_j - x_i)}$$

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

The determinant of A is non-zero if and only if all x_i are distinct.
→ So A is an invertible matrix

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)

The determinant of A is non-zero so A is an invertible matrix

$$Aw = y \Rightarrow \underbrace{A^{-1}A}_I w = A^{-1}y \Rightarrow w = A^{-1}y$$

To calculate A^{-1} , we need to use this formula:

$$\rightarrow A^{-1} = \frac{1}{\det(A)} \times \text{adj}(A)$$

with $\text{adj}(A)$ the transpose of the cofactor matrix of A

$$\rightarrow \text{adj}(A) = (-1)^{i+j} \det(A_{i,j}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m))$$

$$\text{adj}(A) = (-1)^{i+j} S_{m-i,j} \prod_{k \neq j}^m (x_k - x_i)$$

$$\text{and } \det(A) = \prod_{0 \leq i < j \leq n} (x_j - x_i)$$

So we have this formula:

$$\rightarrow A^{-1} = \frac{1}{(-1)^{i+j} S_{m-i,j} \prod_{k \neq j}^m (x_k - x_i)} \times \prod_{0 \leq i < j \leq n} (x_j - x_i)$$

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 $Aw = y$?

Suppose that $n > d + 1$ so we cannot compute the inverse of A (not a square matrix)

To solve $Aw = y$ we can use the least squares solution

$$\rightarrow w = (A^t A)^{-1} A^t y$$

with $A^t \Rightarrow$ transpose of matrix A

$(A^t A)^{-1} \Rightarrow$ inverse of product of A^t and A

In order to do that, we need to follow these points

1. Calculate A^t : The transpose of A

2. Compute $A^t A$

3. Compute the inverse of $(A^t A)$

4. Calculate the product of $(A^t A)^{-1}$ and A^t

5. Multiply $(A^t A)^{-1} A^t$ and y

\Rightarrow Now we have $w = (A^t A)^{-1} A^t y$