Polynomial Regression (Handwriting Assignment)

Name: 주한진

Student ID: 20231581

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 nth degree polynomial in x.

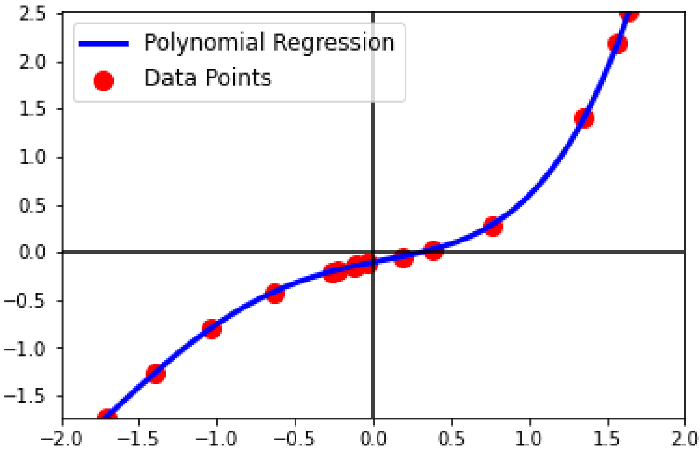


Figure 1: Example of Polynomial Regression

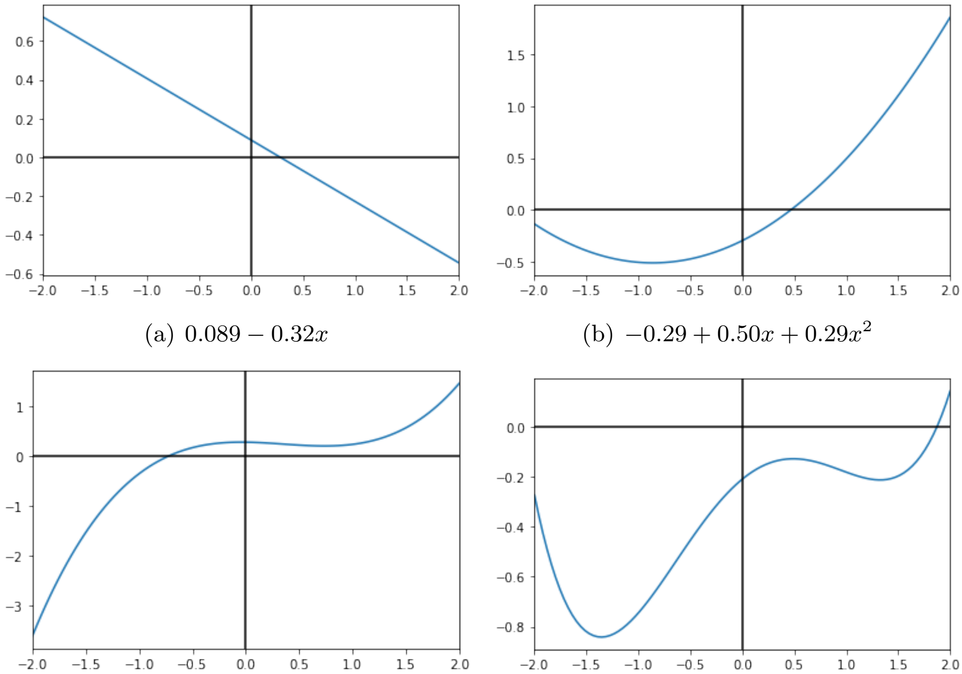
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression 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,

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

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

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

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



(a) $0.089 - 0.32x$ (b) $-0.29 + 0.50x + 0.29x^2$

(c) $0.28 - 0.03x - 0.34x^2 + 0.32x^3$ (d) $-0.21 + 0.34x - 0.35x^2 - 0.06x^3 + 0.1x^4$

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:

$$
\begin{aligned}
\text{Degree of } 0 &: f(x) = w_0 \\
\text{Degree of } 1 &: f(x) = w_1 \cdot x + w_0 \\
\text{Degree of } 2 &: f(x) = w_2 \cdot x^2 + w_1 \cdot x + w_0 \\
\text{Degree of } 3 &: f(x) = w_3 \cdot x^3 + w_2 \cdot x^2 + w_1 \cdot x + w_0 \\
&\vdots \\
\text{Degree of } d &: f(x) = \sum_{i=0}^{d} w_i \cdot x^i,
\end{aligned}
$$

where $w_0, w_1, \cdots, w_d$ 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 $\{w_0, w_1, \cdots, w_d\}$. 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, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Let the degree of polynomial be d. Then, we want to find $w_0, w_1, w_2, \cdots, w_d$ of the polynomial such that

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

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

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

Similarly, we have,

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

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

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

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

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

size ot vector w = $d+1$

size of vector y = n

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

size ot matrix = $n \times (d+1)$

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

1. 행렬식(Determinant) 특징

(1) $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$, $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$

→ 행교환 발생 시, 행렬식의 부호가 바뀜

(2) $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix}$

→ 하나의 row에 대해 나머지 row를 유지한 채 분리 가능

2. Big Formula

Square matrix A가 all-zero 행을 갖는 경우 → $\det A = 0$,

1-(2). 특징에 따라 matrix의 행렬식을 분리하면

분리된 행렬식의 각 열에 1개의 non-zero element가 존재한다.

→ $n \times n$ 의 matrix A의 행렬식을 분리했을 때,

0이 아닌 식은 Permutation matrix 의 행렬식으로, $n!$개 존재한다.

이 중, 절반은 1-(1). 에 의해 +, 나머지는 −의 부호를 갖는다.

이를 식으로 나타내면,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$ 라고 할 때,

$$\det A = \sum_{n!\ \text{terms}} \pm a_{1\alpha} a_{2\beta} \cdots a_{n\omega},$$

$(\alpha, \beta, \cdots \omega) = (1, 2, \cdots n)$의 수열 (= permutation)

3. Big Formula 대입

위 Big Formula 식에

$$A = \begin{bmatrix} 1 & a_1^1 & \cdots & a_1^{n-1} \\ 1 & a_2^1 & \cdots & a_2^{n-1} \\ & & \vdots & \\ 1 & a_n^1 & \cdots & a_n^{n-1} \end{bmatrix}$$

를 대입하면,

$$\det A = \sum_{n!\ \text{terms}} a_1^{\alpha} a_2^{\beta} \cdots a_n^{w}$$

$(\alpha, \beta, \cdots w) = (0, 1, \cdots n-1)$의 순열 로 $\det A$를 표현할 수 있다.

$$\therefore \det A = \sum_{n!\ \text{terms}} a_1^{\alpha} a_2^{\beta} \cdots a_3^{w},$$

$(\alpha, \beta, \cdots w) = (0, 1, \cdots n-1)$의 순열 //

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

If $a_1, a_2, \cdots, a_n$ all have the same value,

A has equal row, which means, $\det A = 0$

$\therefore$ $a_1, a_2, \cdots a_n$ has at least one different value

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

$Aw = y$

$A^{-1}Aw = A^{-1}y$ $\quad A^{-1}A = I$

$w = A^{-1}y$

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

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

$Aw = y,$

$\downarrow$ change into square matrix

$A^T A w = A^T y,$

$(A^T A)^{-1}(A^T A) w = (A^T A)^{-1} A^T y$

$\downarrow$ $(A^T A)^{-1}(A^T A) = I$

$w = (A^T A)^{-1} A^T y$

$\downarrow$ $(A^T A)^{-1} A^T =$ pseudo inverse of A

$w = A^{+} y$

$\therefore \underline{w = A^{+} y}$