Polynomial Regression (Handwriting Assignment)

Name: 이종원

Student ID: 20235859

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.

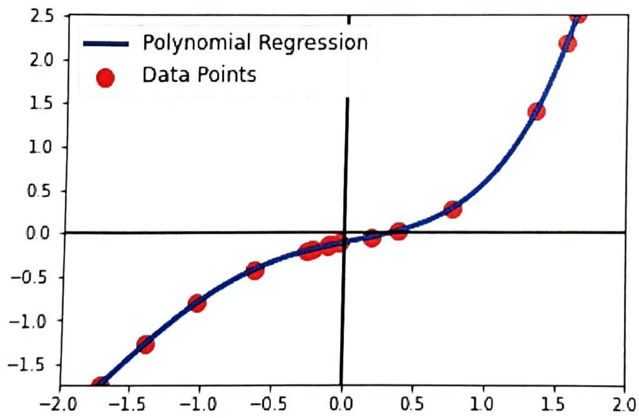


Figure 1: Example of Polynomial Regression

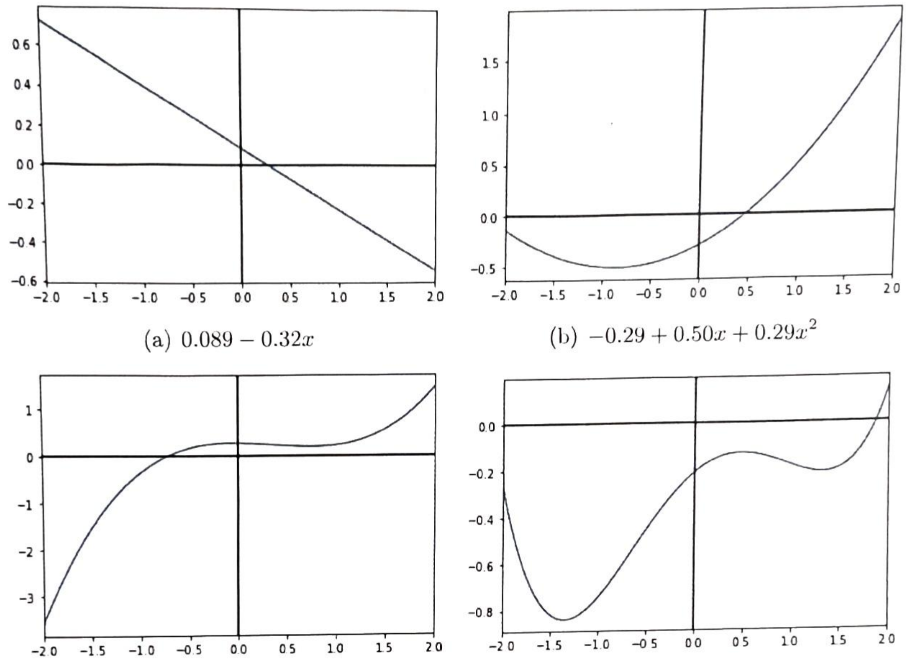
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

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

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

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

Now, we reformulate the equations into the vector and matrix form. First, let $\mathbf{w} = [w_0, w_1, \cdots, w_d]^T$ 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)

$$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

size of vector $w = d+1$

size of vector $Y = n$

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

$$A = \overbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^d \\ & & & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^d \end{bmatrix}}^{d+1} \Big\} n$$

size of matrix A = $\underline{n \times (d+1)}$ matrix

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

$$A = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^d \\ & & & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^d \end{bmatrix}}_{d+1} \Big\} n$$

$d+1=n \Rightarrow d=n-1$

A is $n \times n$ square matrix

$$\det A = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ & & \vdots & & \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix} = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 0 & x_2-x_1 & x_2^2-x_1^2 & \cdots & x_2^{n-1}-x_1^{n-1} \\ 0 & x_3-x_1 & x_3^2-x_1^2 & \cdots & x_3^{n-1}-x_1^{n-1} \\ & & \vdots & & \\ 0 & x_n-x_1 & x_n^2-x_1^2 & \cdots & x_n^{n-1}-x_1^{n-1} \end{vmatrix}$$

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

$j \in \{1,2,3,\ldots,n-1\}$에 대하여 $j+1$ 열에서 j항의 x_1배를 빼는 기본열 연산을 내림차순으로

$$= \prod_{i=2}^{n}(x_i - x_1) \underline{\begin{vmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ 1 & x_3 & \cdots & x_3^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} \end{vmatrix}}$$

→ 방데르몽드 행렬이 다시 등장하기 때문에 반복

$$\det A = \prod_{k=1}^{n-1}\left(\prod_{i=k+1}^{n}(x_i - x_k)\right) = \prod_{1 \le j < i \le n}(x_i - x_j)$$ 을 얻을 수 있다

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

$\det A = \prod_{1 \le j < i \le n} (x_i - x_j) \neq 0$ 이 성립하기 위해서는 $(x_i - x_j)$ 에서

0 이 만들어지는 경우가 배제되어야 한다.

$\forall_i \forall_j \, (x_i \neq x_j \, , \; i \neq j)$ 일때 determinant of A가 non-zero가 된다

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)

$\det A \neq 0$ 이라는 것은 n×n 행렬 A의 역행렬이 존재한다는 것이다

$Aw = y, \; A^{-1}Aw = A^{-1}y \Rightarrow Iw = A^{-1}y$

이때 I 는 단위행렬이므로 $w = A^{-1}y$

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

$n > d+1$ 인 경우 A의 역행렬을 구할 수 없기 때문에

A에 대한 의사 역행렬, A^+를 가정하여 해를 구한다.

이때 의사 역행렬 A^+는 (Pseudo Inverse Matrix)

$AA^+A = A$ 와 $A^+AA^+ = A^+$ 를 만족해야 한다.

이때 $A^+ = (A^TA)^{-1}A^T$ 라고 하는데

이는 A가 선형독립 열을 가져 A^TA 의 역행렬이 존재하는 상황에서

증명이 가능하다. 처음식 $Aw = y$를 통해 증명해 보이면

$Aw = y$, $A^TAw = A^Ty$ (이때 A^TA는 $n \times n$ 행렬)

$$(A^TA)^{-1}(A^TA)w = (A^TA)^{-1}A^Ty$$

$$Iw = (A^TA)^{-1}A^Ty$$

$w = \underline{(A^TA)^{-1}}A^Ty$ 가 성립함을 알 수 있다

즉 $w = A^+y$ $(A^+ = (A^TA)^{-1}A^T)$