Polynomial Regression (Handwriting Assignment)

Name: 박경서

Student ID: 20206622

Instructor: Professor Kyungjae Lee

October 21, 2021

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.

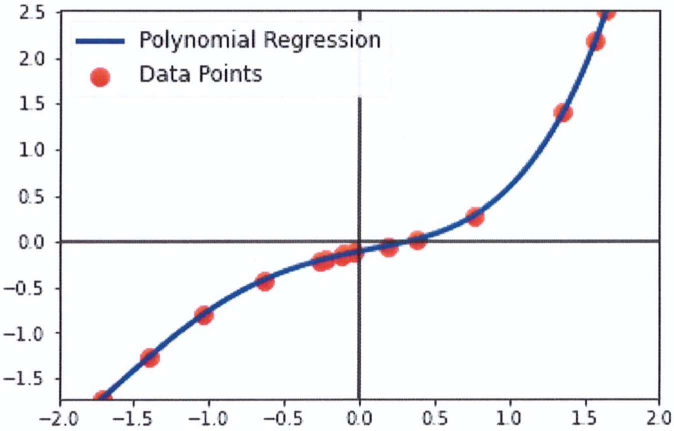


Figure 1: Example of Polynomial Regression

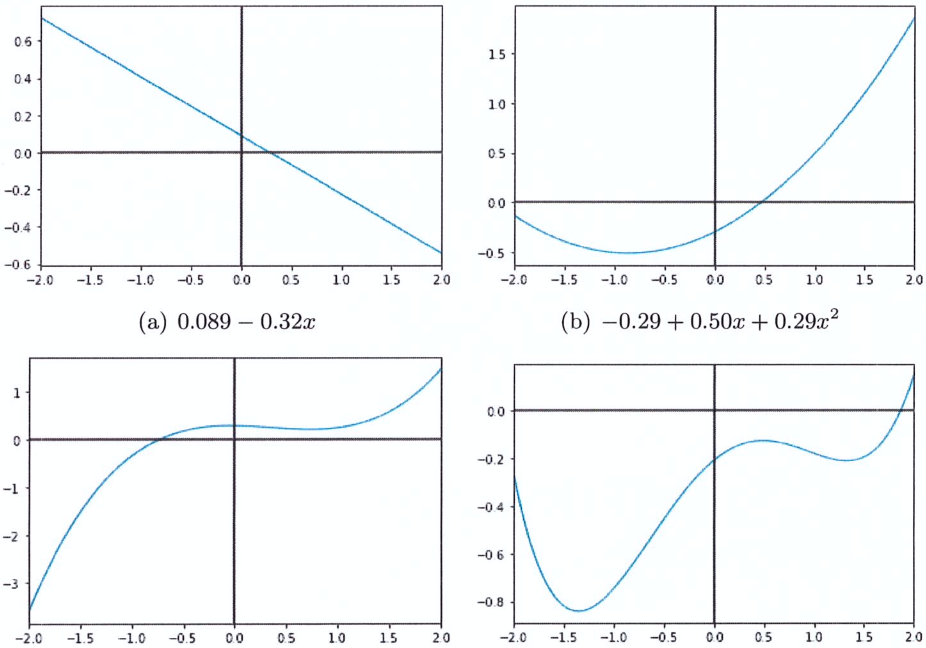
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

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

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)

w는 원소 d+1개로 구성된 열벡터이므로, w의 크기는 d+1이다.

y는 원소 n개로 구성된 열벡터이므로, y의 크기는 n이다.

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

$$A = \begin{bmatrix} 1, x_1, x_1^2, x_1^3, x_1^4 \cdots x_1^d \\ 1, x_2, x_2^2, x_2^3, x_2^4 \cdots x_2^d \\ \vdots \\ 1, x_n, x_n^2, x_n^3, x_n^4 \cdots x_n^d \end{bmatrix}$$

A는 $n \times (d+1)$ 행렬이므로, A의 크기는 $n(d+1)$이다.

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

1. $n \times n$ vandermonde 행렬 A_n의 행렬식 $\det(A_n)$은 다음과 같이 정의된다.

$$\det(A_n) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

2. 명제 $P(n) : \det(A_n) = \prod_{1 \le i < j \le n} (x_j - x_i)$에 대하여

① $n=2$일 때, $\det(A_2) = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$ 이므로 $P(2)$는 참이다.

② $k \ge 3$인 자연수 k에 대해 $P(k)$가 참이라고 가정하고, $P(k+1)$에 대해

$$\det(A_{k+1}) = \begin{vmatrix} 1 & x_1, & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & & \vdots & & \\ 1 & x_{k+1} & x_{k+1}^2 & \cdots & x_{k+1}^k \end{vmatrix}$$

이다. 여기에 1행을 제외한 다른 모든행에서 1행을 빼는 기본행연산을 수행하면

$$\det(A_{k+1}) = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & \cdots & x_2^k - x_1^k \\ \vdots & \vdots & & & \\ 0 & x_{k+1} - x_1 & \cdots & & x_{k+1}^k - x_1^k \end{vmatrix}$$

을 얻는다. 여기에 1열을 제외한 각 열 $R_m (m=2,3,\dots k)$에서 R_{m-1}의 x_1배를 빼는 열연산을 수행하면

$$\det(A_{k+1}) = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & x_2(x_2 - x_1) & \cdots & x_2^{k-1}(x_2 - x_1) \\ \vdots & \vdots & \vdots & & \\ 0 & x_{k+1} - x_1 & x_{k+1}(x_{k+1} - x_1) & \cdots & x_{k+1}^{k-1}(x_{k+1} - x_1) \end{vmatrix}$$

을 얻는다.

이 때, 위의 행렬식은 $\left(\prod_{i=2}^{k+1}(x_i - x_1)\right) \begin{vmatrix} 1 & x_2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & & \\ 1 & x_{k+1} & \cdots & x_{k+1}^{k-1} \end{vmatrix} = \left(\prod_{i=2}^{k+1}(x_i - x_1)\right)\left(\prod_{2 \le j < l \le k+1} (x_l - x_j)\right)$로 나타낼 수 있고,

이는 $\prod_{1 \le i < j \le k+1} (x_j - x_i)$과 같으므로, $P(k)$가 참일 때 $P(k+1)$도 참이 된다.

③ ①과 ②에 의하여 명제 $P(n) : \det(A_n) = \prod_{1 \le i < j \le n} (x_j - x_i)$은 2 이상의 자연수 n에 대해 참이다.

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

$x_1 \ne x_2 \ne x_3 \ne \cdots \ne x_n$

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

문제의 조건에 의해, $\det(A) \ne 0$ 이고 A는 $n \times n$ 정방행렬이므로, A의 역행렬 A^{-1}이 존재한다.

따라서 다음과 같이 w를 구할 수 있다.

$Aw = y$

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

$\to Iw = A^{-1}y$ (I는 $n \times n$ 단위행렬)

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

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 $A\mathbf{w} = \mathbf{y}$? (Hint: Pseudo Inverse)

A가 정방행렬이 아닌 경우 A의 의사역행렬을 이용하여 w를 구할 수 있다.

이 때, A의 의사역행렬 A^+는 $A^+ = (A^TA)^{-1}A^T$과 같이 계산되므로,

이를 이용하여 다음과 같이 w를 구할 수 있다.

$$Aw = y$$

$$\rightarrow A^TAw = A^Ty$$

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

$$\rightarrow \quad w = A^+y$$