Polynomial Regression (Handwriting Assignment)

Name: 최은혁

Student ID: 20212451

Instructor: Professor Kyungjae Lee

October 9, 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.

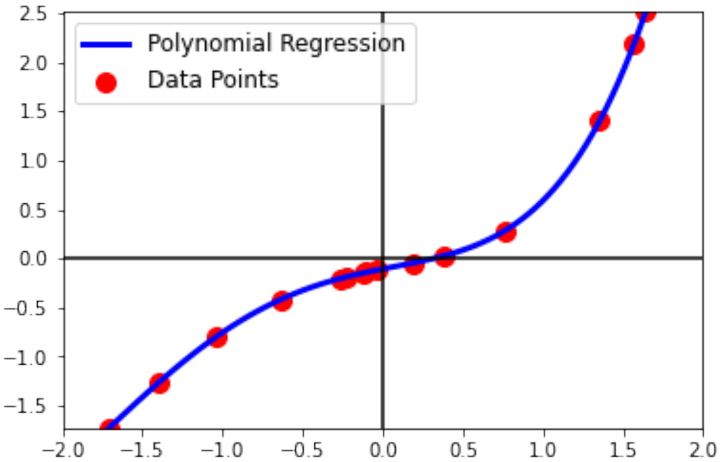


Figure 1: Example of Polynomial Regression

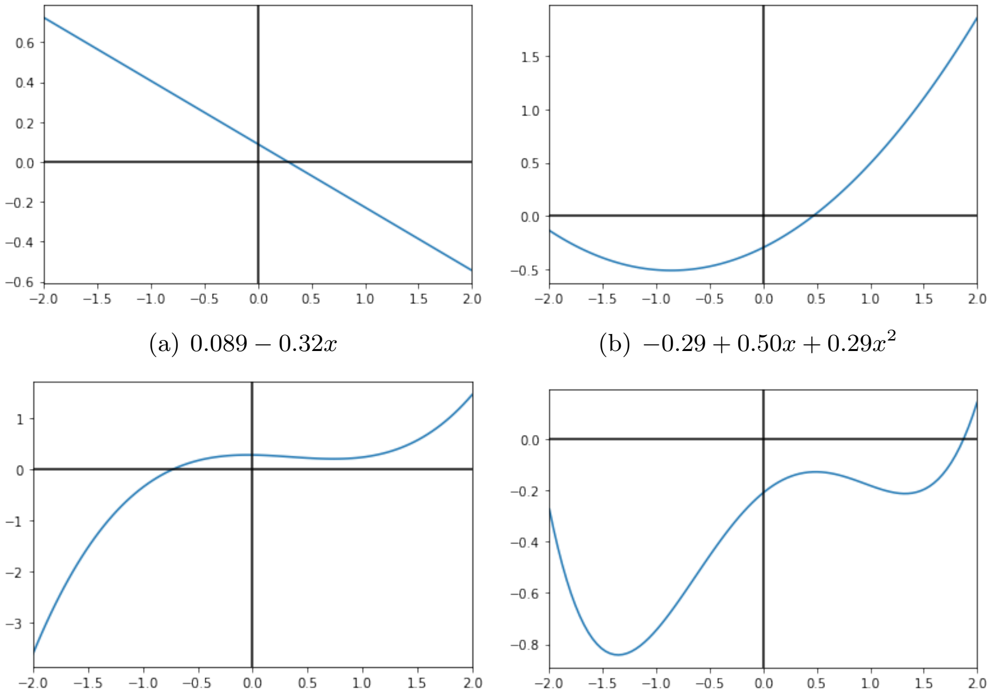
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

$n \times (d+1)$

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

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 : d+1

y : n

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

$A : [n \times (d+1)]$ matrix

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

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

$d+1=n$.

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$: Vandermonde matrix.

$|A| = \prod_{1 \le i < j \le n} (x_j - x_i)$ 라고하자

$n=1$: $|A|_1 = |1| = 1$.

$n=2$: $|A|_2 = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$

$n=1,2$일때 가정 성립.

$n=d$ 일때 $|A|_d = \prod_{1 \le i < j \le d} (x_j - x_i)$ 가 성립한다고 가정.

$n=d+1$: $$A_{d+1} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix}$$

$(d+1)$행을 선택하면 라플라스 전개에 따라

$|A|_{d+1} = 1 \cdot C_{(d+1)1} + x C_{(d+1)2} + x^2 C_{(d+1)3} + \cdots + x^d C_{(d+1)(d+1)}$ 와 같이

d차 다항함수 $f(x)$로 나타낸다. 여기서 x에 $x_1, x_2, \ldots x_d$ 를 대입해보면

완전 같은 행이 여러개 생기므로 $f(x_1) = f(x_2) = \cdots = f(x_d) = 0$

인수정리와 대수학의 기본정리에 의해

$f(x) = C(x-x_1)(x-x_2)\cdots(x-x_d)$ 로 나타낸다

한편 x^d의 계수 $C_{(d+1)(d+1)}$는 $(d+1)$행과

$(d+1)$열에 대한 여인자이므로

$$C_{(d+1)(d+1)} = (-1)^{(d+1)+(d+1)} \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{d-1} \end{vmatrix} = 1 \cdot \prod_{1 \le i < j \le d} (x_j - x_i)$$

다시말해 $C = C_{(d+1)(d+1)}$ 이므로

$$|A|_{d+1} = f(x) = \left(\prod_{1 \le i < j \le d} (x_j - x_i) \right) \left((x-x_1)(x-x_2)(x-x_3)\cdots(x-x_d) \right)$$

$x_{d+1} = x$ 라고 하면 $n=d+1$ 일때 가정이 성립한다.

$\therefore$ 수학적 귀납법에 의해

$|A| = \prod_{1 \le i < j \le d} (x_j - x_i)$ 이다.

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

$|A| = \prod_{1 \le i < j \le d} (x_j - x_i)$ 이므로 모든 항이 0이 아니어야 한다 $\Rightarrow$ $1 \le i < j \le n$, $i \ne j$ 일때 $x_i \ne x_j$.

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)

A는 정사각행렬이고 $|A| \ne 0$ 이므로 A^{-1} 존재한다.

$\therefore$ $W = A^{-1} y$ 를 푼다.

2. (20pt)

Suppose that $n > d$. 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)

$A = [n \times (d+1)]$ matrix

$n > d+1$ 이면 정사각행렬이 아니므로

유사역행렬을 구한다.

유사역행렬 A^+ 는 $AA^+A = A$, $A^+AA^+ = A^+$,

$(AA^+)^T = AA^+$. $(A^+A)^T = A^+A$ 를

만족하면 유사역행렬이라 한다.

$A = U \Sigma V^T$, 3개의 행렬의 곱으로 특이값 분해를 한다.

$U = m \times n$ 직교행렬, $V = n \times n$ 직교행렬이다.

$U^{-1} = U^T$. $V^{-1} = V^T$ 를 만족한다.

$m \times n$ matrix $\Sigma = \begin{bmatrix} \Sigma_r & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{bmatrix}$.

O행렬의 아래첨자는 행과 열의 개수, Σ_r 은 $r \times r$ 대각행렬로

대각성분은 양의 실수이며 내림차순 정렬됨.

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. 이 σ 는 특이값이며 r은 rank A

$A^+ = V \Sigma^+ U^T$

Σ^+ : $n \times m$ matrix, $\Sigma^+ = \begin{bmatrix} \Sigma_r^{-1} & O_{r \times (m-r)} \\ O_{(n-r) \times r} & O_{(n-r) \times (m-r)} \end{bmatrix}$

[A] : $n \times (d+1)$ 에서 $n > d+1$ 이고 rank 가 $r = d+1$ 이라면

$A_L^+ A = I$, $A_L^+ = (A^T A)^{-1} A^T$

이 경우 $A = U \begin{bmatrix} \Sigma_r \\ 0 \end{bmatrix} V^T$

$\therefore A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \end{bmatrix} U^T$: 유사역행렬.

$\therefore w = A^+ y$ 로 계산한다.