Polynomial Regression (Handwriting Assignment)

Name: 이현진

Student ID: 20230520

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.

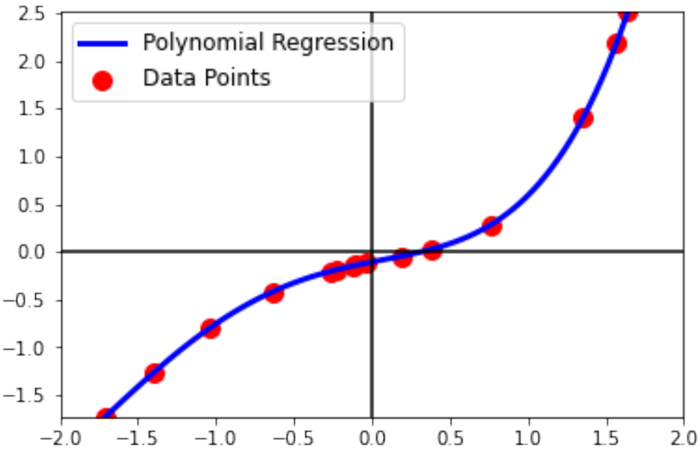


Figure 1: Example of Polynomial Regression

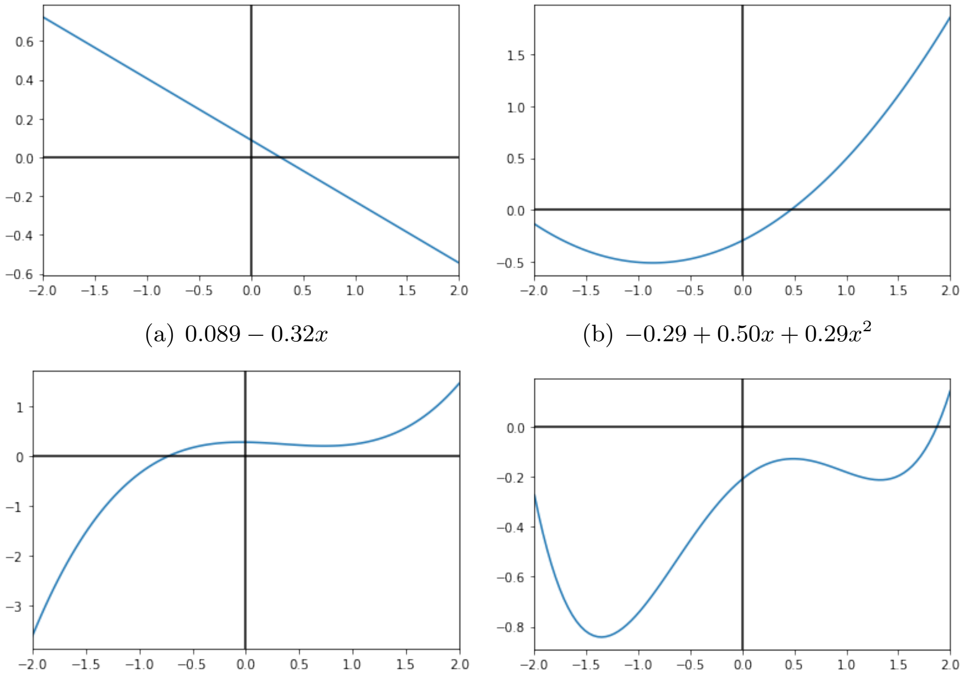
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

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

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)

$w = [w_0, w_1, \cdots, w_d]^T$ 즉 $(1+d, 1)$, $(1+d) \times 1$ 행렬

$|w| = \sqrt{(1+d)^2+1} = \sqrt{2+2d+d^2}$

$y = [y_1, y_2, y_3, \cdots, y_n]^T$ 즉 $(n, 1)$, $n \times 1$ 행렬

$|y| = \sqrt{n^2+1}$

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

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

A의 크기 : $n \times (1+d)$ 행렬

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

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & x_3^3 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{d+1} & x_{d+1}^2 & x_{d+1}^3 & \cdots & x_n^{n-1} \end{pmatrix} = V_n \text{ 라 두자}$$

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

수학적 귀납법을 사용하자 $P(n) : \det(V_n) = \prod_{1\le i \le j \le n} (x_j - x_i)$ 라 정의하자

$P(1) = \det V_1 = |1| = 1$

$P(2) = \det V_2 = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$

따라서 $n=1,2$일 때 $P(n)$은 성립

$n=k$; $P(k) : \det V_k = \prod_{1\le i \le j \le k} (x_j - x_i)$ 가 성립한다 가정

$n=k+1$;

$$V_k' = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & & & & \\ 1 & x_k & x_k^2 & \cdots & x_k^k \\ 1 & x & x^2 & \cdots & x^k \end{bmatrix},$$

$(k+1)$행을 선택하여 라플라스 전개를 쓰면

$$|V_k'| = 1\, C_{(k+1)1} + x\, C_{(k+1)2} + x^2 C_{(k+1)3} + \cdots + x^k C_{(k+1)(k+1)} = f(x) \text{ 라 두면}$$

여기에 $x = x_1 \cdots x_k$ 를 대입하면 $|V_k'|$에 완전히 같은 두 행이 생기므로 0이 됨.

$\therefore f(x) = C(x-x_1)(x-x_2)\cdots(x-x_k)$ ($\because$ 인수정리)

C는 $(k+1)$행 $(k+1)$열에 대한 여인자이므로

$$C = (-1)^{(k+1)+(k+1)} \det(V_{(k)}) = 1 \cdot \prod_{1\le i \le j \le k} (x_j - x_i)$$

$\therefore \det(V_k') = f(x) = \left[\prod_{1\le i\le j\le k}(x_j - x_i)\right]((x-x_1)(x-x_2)\cdots(x-x_k))$, $x = x_{k+1}$ 라 정의하면 $V_k' = V_{k+1}$,

$P(k+1) = \det V_{k+1} = \prod_{1\le i \le j \le k+1} (x_j - x_i)$ 따라서 $P(k)$가 참일때 $P(k+1)$이 참이므로 수학적귀납법에 의해 증명됨.

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

① 모든 행,열 벡터들이 선형적으로 독립이어야 한다.

② 영벡터가 행이나 열 벡터 중 아무곳에서도 존재하면 안된다.

③ 행렬의 랭크가 $n \times n$ 행렬이라면 n, 즉 최대여야 한다.

$n \times n$ 정방행렬 A의 행렬식의 값이 0이 아닐때,

A는 정칙행렬이다. A가 정칙임과 $|A| \neq 0$ 은 서로 필요충분조건이기 때문.

정칙행렬은 가역적이기에 역행렬이 존재한다.

즉, $|A| \neq 0$ 이면 $AB = BA = I$ 를 만족하는 A의 역행렬 $A^{-1} = B$ 가 존재한다.

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)

$Aw = y$ Use normal equation

손실함수 $L(w) = (y - Aw)^T (y - Aw) = \|y - Aw\|^2$

$$\frac{\partial}{\partial w} L(w) = \frac{\partial}{\partial w} \left((y - Aw)^T (y - Aw) \right)$$

$$= \frac{\partial}{\partial w} \left(y^T y - y^T A w - w^T A^T y + w^T A^T A w \right)$$

$$= 0 - A^T y - A^T y + 2 A^T A w$$

$\therefore L(w) = 0$ 일 때

식을 정리하면 $\Rightarrow A^T A w = A^T y$

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

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

비정방행렬 → 유사역행렬을 구해야 함

특이값 분해 후 최소자승법 이용

$A = USV^T$ $\quad\quad UU^T = I_{n\times n}$

$\quad\quad\quad\quad\quad\quad VV^T = I_{p\times p}$

$AA^T = USV^T \cdot (USV^T)^T = USV^T \cdot V S^T U^T$

$= US(V^TV)S^TU^T = USS^TU^T$

$= UD_uU^T \quad\quad \Lambda = S^2 \mapsto$ A의 0이 아닌 특이값들은 A^TA 또는 AA^T의 고유값들의 제곱근과 동일

(∴ U는 AA^T의 고유벡터로 이루어진 행렬

V는 A^TA의 고유벡터로 이루어진 행렬 ← 고유값 분해로 찾는다.

① $\det(AA^T - \lambda I) = 0$ 인 λ값들 탐색

② $(AA^T - \lambda I)\cdot u = 0$ 인 고유벡터 V 탐색

③ 고유값 분해 $AA^T = uDu^{-1}$

유사역행렬 $A^+ = VS^+U^T$ $\quad$ S^+는 S의 역수를 취한 대각행렬,

0이 아닌 특이값들을 역수로 취함.

ex) $S = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad$ $S^+ = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & 0 \end{pmatrix}$

$(m \times n) \quad\quad\quad\quad (n \times m)$

$w = A^+y$

$= (VS^+U^T)y$