

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 n th degree polynomial in x .

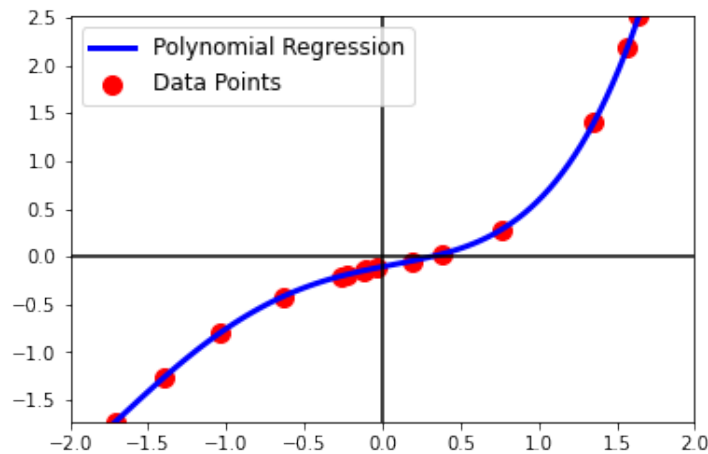


Figure 1: Example of Polynomial Regression

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,

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

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

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

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



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:

$$\text{Degree of 0 : } f(x) = w_0$$

$$\text{Degree of 1 : } f(x) = w_1 \cdot x + w_0$$

$$\text{Degree of 2 : } f(x) = w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\text{Degree of 3 : } f(x) = w_3 \cdot x^3 + w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\vdots$$

$$\text{Degree of } d : f(x) = \sum_{i=0}^d w_i \cdot x^i,$$

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

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

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

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

Similarly, we have,

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

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

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

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

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

$$w = [w_0, w_1, \dots, w_d]^T \quad \approx (1+d, 1), (d+1) \times 1 \text{ 행렬}$$

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

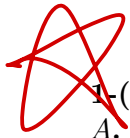
$$y = [y_1, y_2, y_3, \dots, y_n]^T \quad \approx (n, 1), n \times 1 \text{ 행렬}$$

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

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

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

A 의 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)

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^d \\ 1 & x_3 & x_3^2 & x_3^3 & \dots & x_3^d \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{d+1} & x_{d+1}^2 & x_{d+1}^3 & \dots & x_{d+1}^d \end{pmatrix}$$

$$\det(A) = \sum_{k=1}^{d+1} (-1)^{k+1} \cdot 1 \cdot \mu_{k1} \quad (\because \text{라플라스 전개})$$

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 A 의 행렬식의 값이 0이 아닐때,

A 는 정칙행렬이다. A 가 정칙일과 $|A| \neq 0$ 은 서로 필요충분조건이기 때문이다.

정칙행렬은 가역적이기에 역행렬이 존재한다.

즉, $|A| \neq 0$ 이면 $AB = BA = I$ 를 만족하는 A 의 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, $Aw = y$, with respect to w ? (10pt)

$Aw = y$ Use normal equation

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

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

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

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

$$\therefore L(w) = 0 \text{ 일 때}$$

$$\text{식을 정리하면} \Rightarrow A^T A w = A^T y$$

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

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 \rightarrow 유사역행렬을 구해야 함

특이값 분해 후 최소자승법 이용

$$A = U S V^T \quad \begin{aligned} U U^T &= I_{n \times n} \\ V V^T &= I_{p \times p} \end{aligned}$$

$$A A^T = U S V^T (U S V^T)^T = U S V^T V S^T U^T$$

$$= U (S^T V^T V) S^T U^T = U S S^T U^T$$

$$= U \Lambda U^T$$

$\Lambda = S^2 \rightarrow A$ 의 0이 아닌
특이값들은

$A^T A$ 또는 $A A^T$ 의 고유값들
의 제곱근과 동일

$\therefore U$ 는 $A A^T$ 의 고유벡터로 이루어진 행렬

V 는 $A^T A$ 의 고유벡터로 이루어진 행렬 \leftarrow 고유값 분해로 찾는다.

$$\textcircled{1} \det(A A^T - \lambda I) = 0 \text{ 인 } \lambda \text{ 값들 탐색}$$

$$\textcircled{2} (A A^T - \lambda I) \cdot u = 0 \text{ 인 고유벡터 } u \text{ 탐색}$$

$$\textcircled{3} \text{ 고유값 분해 } A A^T = U \Lambda U^T$$

$$\text{유사역행렬 } A^+ = V S^+ U^T$$

S^+ 는 S 의 역수를 취한 대각행렬,

0이 아닌 특이값들을 역수로 취함.

$$\mathbf{w} = A^+ \mathbf{y}$$

$$= (V S^+ U^T) \mathbf{y}$$

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

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