Polynomial Regression (Handwriting Assignment)

Name: 경순범

Student ID: 2021 3158

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.

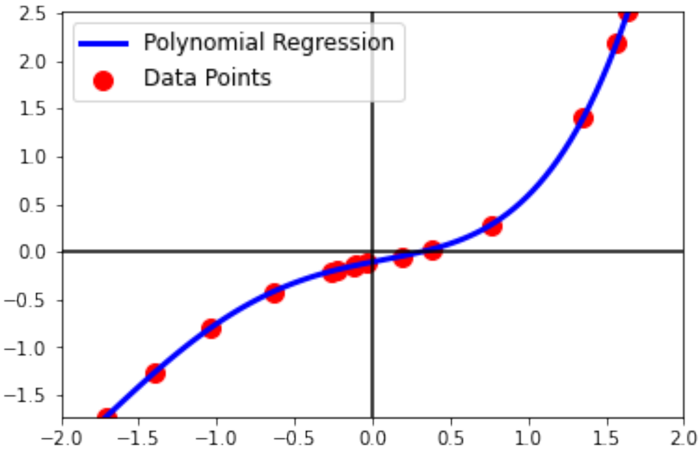


Figure 1: Example of Polynomial Regression

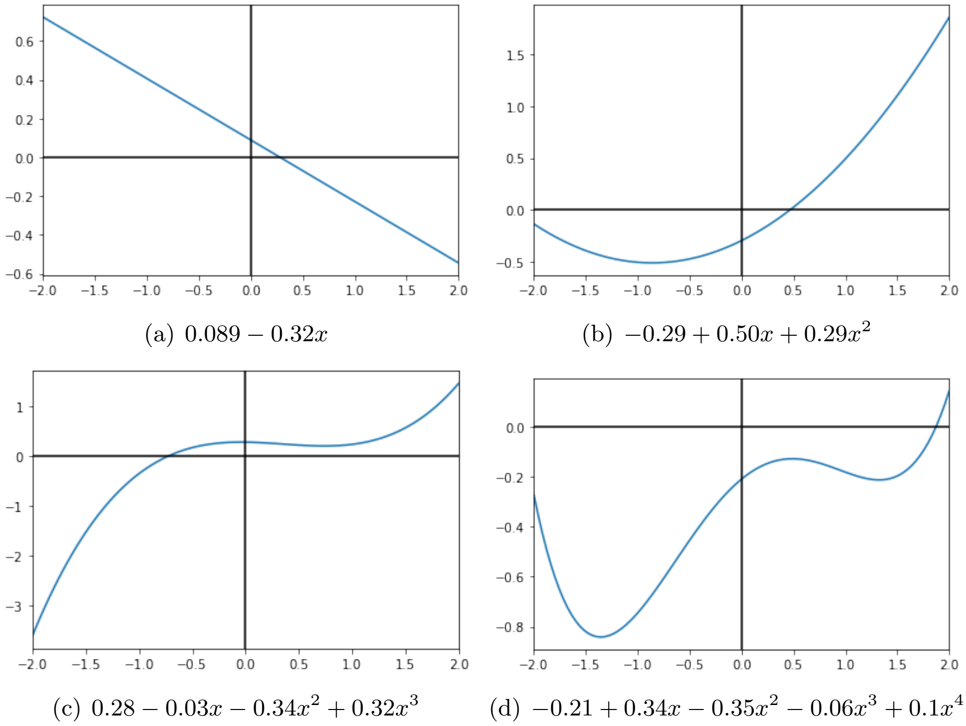
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

1.-(a) 벡터 w는 d+1 개의 요소로 이루어져 있으므로 크기가 d+1이고
벡터 y는 n개의 요소로 이루어져 있으므로 크기가 n 입니다

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

1-(b)

$$\text{matrix } A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^d \\ & & & \vdots & & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^d \end{bmatrix}$$

size of matrix A is 가로의 요소수와 세로의 요소수를 곱한 $n\times(d+1)$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-(c) 예를들어 3x3 matrix를 이용하여 구하고 이것을 nxn으로 확장하겠습니다

$$\det\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{bmatrix} = \det\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 0 & \lambda_2-\lambda_1 & \lambda_2^2-\lambda_1^2 \\ 0 & \lambda_3-\lambda_2 & \lambda_3^2-\lambda_1^2 \end{bmatrix}$$ 에서 추출하면

$$= (\lambda_2-\lambda_1)(\lambda_3-\lambda_1)\det\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 0 & 1 & \lambda_1+\lambda_2 \\ 0 & 1 & \lambda_1+\lambda_3 \end{bmatrix}$$

여기서 2번째 row에 λ_1을 곱하고 1번 row를 빼고
3번째 row에 λ_1을 곱하고 1번 row를 빼면

$$= (\lambda_2-\lambda_1)(\lambda_3-\lambda_1)\det\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \lambda_2 \\ 0 & 1 & \lambda_3 \end{bmatrix}$$ 이것을 정리하면

$$= (\lambda_2-\lambda_1)(\lambda_3-\lambda_1)\det\begin{bmatrix} 1 & \lambda_2 \\ 1 & \lambda_3 \end{bmatrix}$$ 이다 이것은 $\prod_{1\le i<j\le 3}(\lambda_j-\lambda_i) \neq 0$ 이다

이것을 미친것지로 Vandermonte Matrix에 적용하면

$$\det\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \vdots \\ \vdots & & & & \vdots \\ 1 & \lambda_{n-1} & \lambda_{n-1}^2 & \cdots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix}$$ 이것을 3x3 matrix와의 같은방법으로 정리하면

$$\det\begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-3} & \lambda_1^{n-2} & \lambda_1^{n-1} \\ 0 & \lambda_2-\lambda_1 & \cdots & & \lambda_2^{n-2}-\lambda_1^{n-2} & \lambda_2^{n-1}-\lambda_1^{n-1} \\ \vdots & & & & & \\ 0 & \lambda_n-\lambda_1 & \cdots & & \lambda_n^{n-2}-\lambda_1^{n-2} & \lambda_n^{n-1}-\lambda_1^{n-1} \end{bmatrix}$$

$$= \prod_{2\le j\le n}(\lambda_j-\lambda_1)\det\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & & \lambda_n^{n-2} \\ \vdots & \vdots & & & \vdots \\ 0 & 1 & \cdots & & \lambda_n^{n-2} \end{bmatrix}$$ 이것을 정리하면

$$= \prod_{1\le i<j\le n}(\lambda_j-\lambda_i)$$ 임을 알수있습니다

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

1-(d) det A가 0이되면 안되므로 $(\lambda_j-\lambda_i)$ $1\le i<j\le n$ 가 0이아닐때 non zero가 됩니다

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)

1-(e) det A가 0이아니다는 말은 unique solution을 갖거나 inverse가 존재한다는 의미이므로
$Aw=y$ 에서 양변에 A^{-1}를 취하면

$A^{-1}Aw = A^{-1}Y$

$w = A^{-1}Y$ 이므로

답은 $w = A^{-1}Y$ 입니다

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)

DATE.

2. A가 Square matrix가 아닌경우 A inverse를 구할수 없으므로
Pseudo inverse라는 개념을 도입한다
앞시간에 배운 A^+를 수도 inverse 라고하면.
$A^+ = (A^T A)^{-1} A^T$ 이므로
$Aw = y$ 에 양변에 A^T를 곱하면

$A^T A w = A^T y$ 이고

$(A^T A)^{-1} (A^T A) w = (A^T A)^{-1} A^T y$
$I w = (A^T A)^{-1} A^T y$
$\therefore w = (A^T A)^{-1} A^T y$
$\therefore w = A^+ y$ 임을 알수있다