Polynomial Regression (Handwriting Assignment)

Name: 정의연

Student ID: 20220742

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.

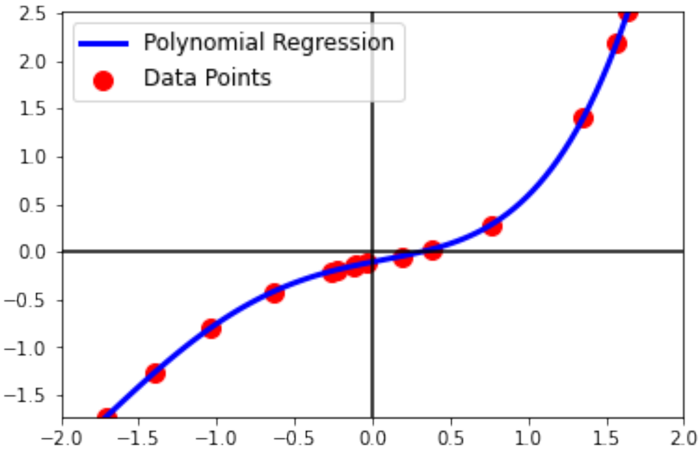


Figure 1: Example of Polynomial Regression

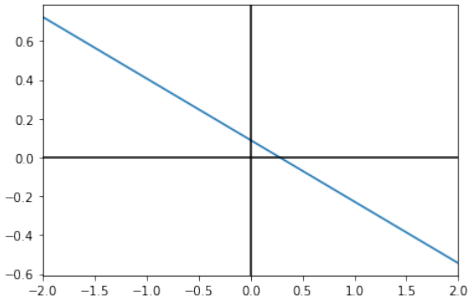
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

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

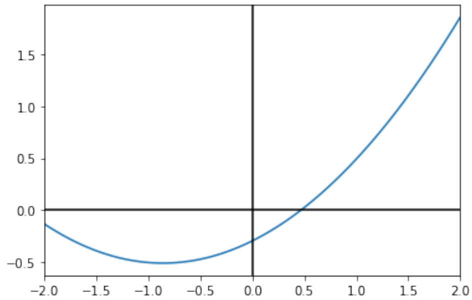
The goal of regression is to find a function $\hat{f}$ such that

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

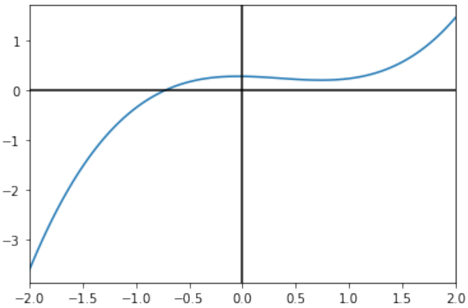
This is the simplest definition of the regression problem. Note that many details about regression analysis are omitted here, but, you will learn more rigorous definition in other courses such as



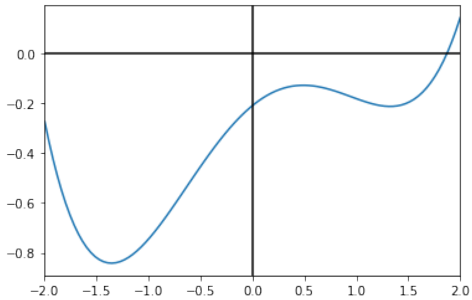
(a) $0.089 - 0.32x$



(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,

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

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)

the size of vector w is d+1 and

the size of vector y is n

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

The size of matrix A is $n \times d+1$

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

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

A_n이 $n \times n$의 square matrix라고 하면 $d = n-1$ 이된다.

Answer: determinant of A_n is $\prod_{1 \le i < j \le n} (x_j - x_i)$

Derivation : use mathmatical induction

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

basic step : $n = 1, 2$ 일때 성립함을 보인다.

$P(1) = \det A_1 = |1| = 1$ 이되며 성립하고

$P(2) = \det A_2 = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1$ 이므로 성립한다.

induction step $n = k$ 일때 성립한다고 가정하면

$\det A_k = \prod_{1 \le i < j \le k} (x_j - x_i)$ 가 성립하고 이때 $n = k+1$ 일때도 성립함을 보인다.

$n = k+1$ 일때는 행이 하나더 생기기 때문에 그 마지막 행의 변수를 x라고 놓고 그 행을 통해

행렬식을 계산하면 $C_{ij} = (-1)^{i+j} M_{ij}$ 라고하고 (i번째 row와 j번째 column을 없앤 matrix의 determinent를 M_{ij} 라고 하면)

$\det A_{k+1} = 1\, C_{(k+1)1} + x\, C_{(k+1)2} + x^2 C_{(k+1)3} + \cdots + x^k C_{(k+1)(k+1)}$ 그런데 이때 이 행렬식은 k차 다항함수이고

이 함수를 $f(x)$라고 할때 $f(x_1) = f(x_2) = \cdots = f(x_k) = 0$ 이므로 $f(x)$는 $C(x - x_1)(x - x_2) \cdots (x - x_k)$ (C는 상수)로 표현할 수 있다.

이때 C는 $f(x)$의 최고차항의 계수와 같으므로 $C = C_{(k+1)(k+1)}$ 와 같다. $C_{(k+1)(k+1)} = \prod_{1 \le i < j \le k} (x_j - x_i)$ 이므로 ($C_{(k+1)(k+1)} = (-1)^{k+1+k+1} \cdot M_{(k+1)(k+1)}$ 이기 때문)

최종적으로 변수 x가 x_{k+1}라고 하면 $f(x) = \left[\prod_{1 \le i < j \le k} (x_j - x_i)\right](x_{k+1} - x_1)(x_{k+1} - x_2) \cdots (x_{k+1} - x_k) = \prod_{1 \le i < j \le k+1} (x_j - x_i)$

이므로 $\det A_{k+1} = \prod_{1 \le i < j \le k+1} (x_j - x_i)$ 이므로 따라서 determinant of A_n는 $\prod_{1 \le i < j \le n} (x_j - x_i)$이다.

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

$x_1 \neq x_2 \neq x_3 \neq \cdots \neq x_n$ 이어야 A의 determinant가 0이되지 않는다.

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)

A의 column은 linearly independent함과 동시에 square matrix이기 때문에 invertiable하다

따라서 $w = A^{-1}y$로 쓸 수 있다.

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)

$x_1 \neq x_2 \neq x_3 \neq \cdots \neq x_n$ 이기 때문에 A의 column은 linearly independent 하다.

그리고

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

이 대각선(diagonal) 에 대하여 대칭이기 때문에 A^TA는 invertible 하다.

따라서

$$Aw = y$$
$$A^TAw = A^Ty$$
$$w = (A^TA)^{-1}A^Ty \quad \text{이다.}$$

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)