Polynomial Regression (Handwriting Assignment)

Name: 김민성

Student ID: 20222395

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.

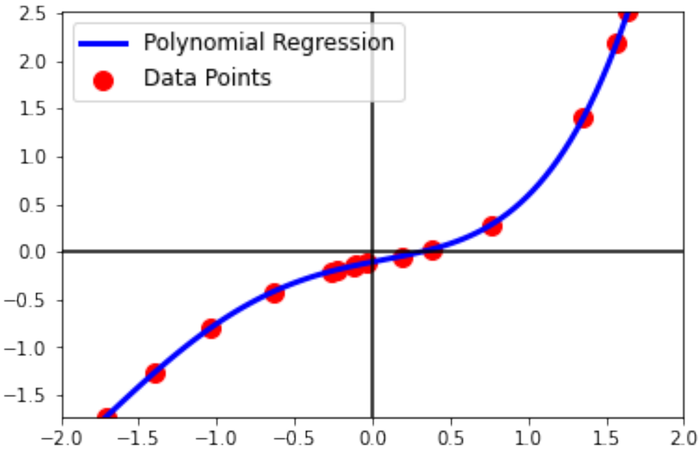


Figure 1: Example of Polynomial Regression

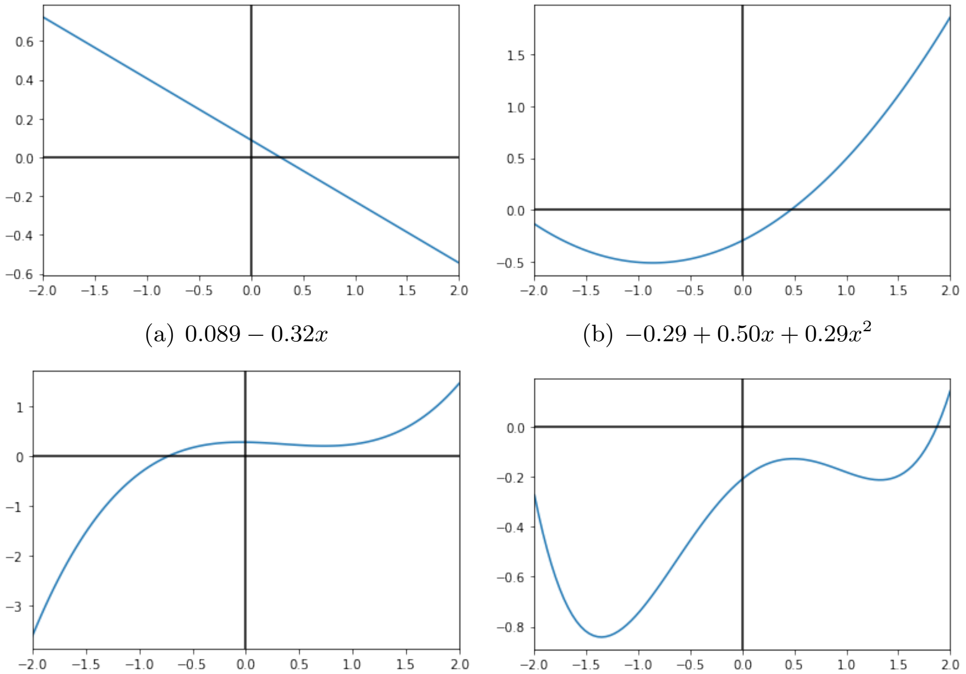
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression 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 $(d+1)$

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 $n \times (d+1)$

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

I'll explain it by using mathematical induction.

Determinant of Vandermonde Matrix is $\prod_{1\le i<j\le n} (x_j - x_i)$

Derivation : Let $P(n) = \prod_{1\le i<j\le n} (x_j - x_i)$ $(n \ge 2)$

$\boxed{\text{Basic step}}$

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}, \quad \det(A) = x_2 - x_1$$

$$P(2) = \prod_{1\le i<j\le 2} (x_j - x_i) = x_2 - x_1 \qquad \therefore P(2) \text{ is true.}$$

$\boxed{\text{Induction step}}$

$$A = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} & x_1^n \\ 1 & x_2 & \cdots & x_2^{n-1} & x_2^n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} & x_n^n \\ 1 & x_{n+1} & \cdots & x_{n+1}^{n-1} & x_{n+1}^n \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & \cdots & x_n & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} & x_{n+1}^{n-1} \\ x_1^n & x_2^n & \cdots & x_n^n & x_{n+1}^n \end{bmatrix} \qquad \det(A) = \det(A^T)$$

In A^T, subtracting x_1 times the i-th row to i+1-th row

$$A^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_{n+1} - x_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_2^n - x_1 x_2^{n-1} & \cdots & x_{n+1}^n - x_1 x_{n+1}^{n-1} \end{bmatrix}$$

Expanding by the first column and factoring $x_i - x_1$ from i-th column $(i = 2, \cdots, n+1)$

$$\det(A^T) = \prod_{j=2}^{n+1} (x_j - x_1) \cdot \det\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ x_2^{n-1} & \cdots & x_{n+1}^{n-1} \end{pmatrix} = \prod_{j=2}^{n+1} (x_j - x_1) \prod_{2\le i<j\le n+1} (x_j - x_i) = \prod_{1\le i<j\le n+1} (x_j - x_i)$$

and the inductive step is complete.

Answer : $\det(A) = \prod_{1\le i<j\le n} (x_j - x_i)$

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

All of x_i $(1 \le i \le n)$ has different value each other.

$\forall x (x_i \ne x_j,\ 1 \le i < j \le n,\ i, j \in \mathbb{N})$

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)

$$Aw = y \iff A^{-1}Aw = A^{-1}y \iff w = A^{-1}y$$

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

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)

Let A^+ be a pseudo inverse of A

From the definition of pseudo inverse,

$$\begin{cases} AA^+A = A \\ A^+AA^+ = A^+ \end{cases}$$

A has linearly independent columns and thus matrix A^TA is invertible.

- $A^+ = (A^TA)^{-1}A^T$

$$y = Aw$$

$$A^Ty = A^TAw$$

$$(A^TA)^{-1}A^Ty = w$$

$$A^+y = w$$

$$\therefore w = (A^TA)^{-1}A^Ty$$