Polynomial Regression (Handwriting Assignment)

Name: 강지헌

Student ID: 20221186

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.

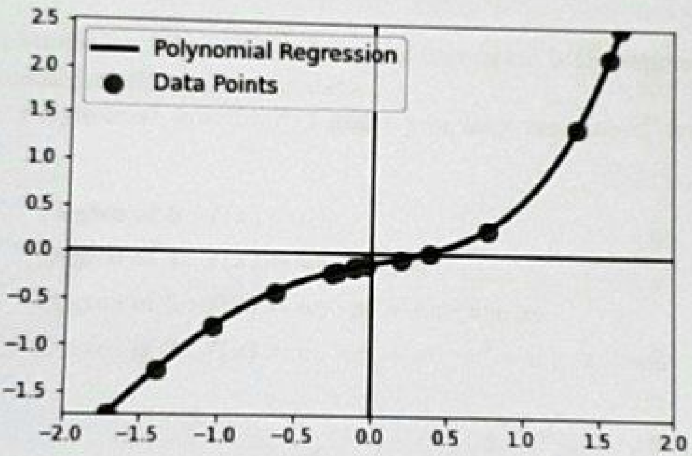

Figure 1: Example of Polynomial Regression

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

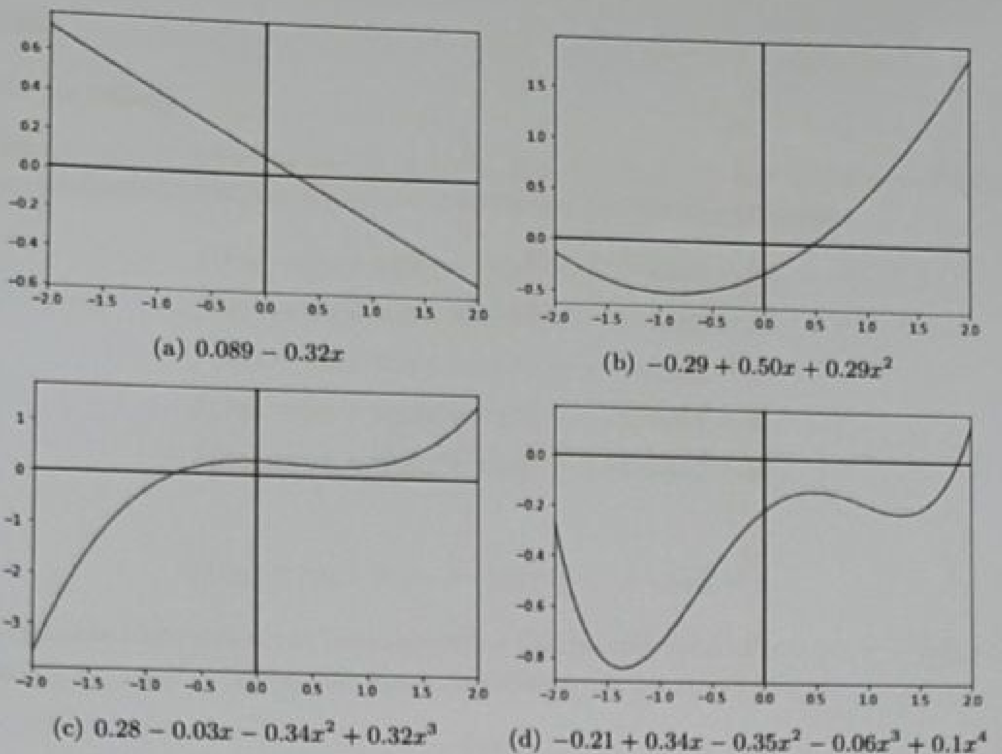

Figure 2: Examples of polynomial functions

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:

$$
\begin{aligned}
&\text{Degree of } 0: f(x) = w_0 \\
&\text{Degree of } 1: f(x) = w_1 \cdot x + w_0 \\
&\text{Degree of } 2: f(x) = w_2 \cdot x^2 + w_1 \cdot x + w_0 \\
&\text{Degree of } 3: f(x) = w_3 \cdot x^3 + w_2 \cdot x^2 + w_1 \cdot x + w_0 \\
&\qquad\vdots \\
&\text{Degree of } d: f(x) = \sum_{i=0}^{d} w_i \cdot x^i,
\end{aligned}
$$

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)

Vector size means how many values the vector has. So size of vector w is d+1, of vector y is n.

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

In linear equation $Aw = y$, w is $[w_0, w_1, \cdots, w_d]^T$ and y is $[y_1, y_2, \cdots, y_n]^T$,

A must be
$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ & & \vdots & & \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix}$$
. So 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.)

when $d = n-1$, A :
$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ & & \vdots & & \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$
. The form of A is Vandermonde matrix.

By row operations,
$$\det A = \begin{bmatrix} 1 & x_1 & x_1^2 \cdots & x_1^{n-1} \\ 0 & x_2-x_1 & x_2^2-x_1^2 \cdots & x_2^{n-1}-x_1^{n-1} \\ & & \vdots & \\ 0 & x_n-x_1 & x_n^2-x_1^2 \cdots & x_n^{n-1}-x_1^{n-1} \end{bmatrix} = \det \begin{bmatrix} x_2-x_1 & x_2^2-x_1^2 & \cdots & x_2^{n-1}-x_1^{n-1} \\ x_3-x_1 & x_3^2-x_1^2 & \cdots & x_3^{n-1}-x_1^{n-1} \\ & & \vdots & \\ x_n-x_1 & x_n^2-x_1^2 & \cdots & x_n^{n-1}-x_1^{n-1} \end{bmatrix}$$
$$= \det\left(\begin{bmatrix} x_2-x_1 & 0 & \cdots & 0 \\ 0 & x_3-x_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n-x_1 \end{bmatrix}\begin{bmatrix} 1 & x_2+x_1 & \cdots & \sum_{i=0}^{n-2} x_2^{n-2-i}\cdot x_1^i \\ 1 & x_3+x_1 & \cdots & \sum_{i=0}^{n-2} x_3^{n-2-i}\cdot x_1^i \\ \vdots & \vdots & & \vdots \\ 1 & x_n+x_1 & \cdots & \sum_{i=0}^{n-2} x_n^{n-2-i}\cdot x_1^i \end{bmatrix}\right)$$

and by Cauchy-Binet Theorem, we can get
$$\det A = \cdots = \prod_{j=2}^{n}(x_j-x_1)\det\begin{bmatrix} 1 & x_2+x_1 & \cdots & \sum_{i=0}^{n-2} x_2^{n-2-i}\cdot x_1^i \\ \vdots & \vdots & & \vdots \\ 1 & x_n+x_1 & & \sum_{i=0}^{n-2} x_n^{n-2-i}\cdot x_1^i \end{bmatrix}$$
$$= \prod_{j=2}^{n}(x_j-x_1)\det\left(\begin{bmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ 1 & x_3 & \cdots & x_3^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} \end{bmatrix}\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-2} \\ 0 & 1 & x_1 & & x_1^{n-3} \\ 0 & 0 & 1 & & x_1^{n-4} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & 1 \end{bmatrix}\right) = \prod_{j=2}^{n}(x_j-x_1)\det\begin{bmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ 1 & x_3 & \cdots & x_3^{n-2} \\ & & \vdots & \\ 1 & x_n & \cdots & x_n^{n-2} \end{bmatrix}$$
. $\left(\prod_{j=2}^{n}(x_j-x_1) = (x_2-x_1)(x_3-x_1)\times\cdots\times(x_n-x_1)\right)$

By repeating this process, we get
$$\det A = \prod_{j=2}^{n}(x_j-x_1)\times\prod_{2\le i<j\le n}(x_j-x_i) = \prod_{1\le i<j\le n}(x_j-x_i).$$

Thus, the determinant of A is $\prod_{1\le i<j\le n}(x_j-x_i)$.

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

The determinant of A is $(x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)(x_3-x_2)\cdots(x_n-x_2)\cdots\cdots(x_n-x_{n-1})$, so when the determinant of A is not zero, the condition must be $x_p \neq x_q$ with premise that $p \neq q$ and $1 \le p, q \le n$.

1-(e) Assume that the determinant of A is non-zero, then, what is the solution of linear equation, $A\mathbf{w} = \mathbf{y}$, with respect to $\mathbf{w}$? (10pt)

Since $\det(A) \neq 0$, and A is square matrix, A has a inverse matrix. So $w = A^{-1}y$, and let $A^{-1} = [a_{ij}]$ $(1 \le i, j \le n)$.

Using the definition of the matrix product and the inverse, $\sum_{k=1}^{n} a_{kj} x_i^{k-1} = \delta_{ij}$ (this means if $i \neq j$, value is 0, if $i = j$, value is 1)

That is, if $P_j(x)$ is the polynomial, let $P_j(x) = \sum_{k=1}^{n} a_{kj} x^{k-1}$. Then $P_j(x_1) = 0, \cdots P_j(x_{j-1}) = 0, P_j(x_j) = 1, P_j(x_{j+1}) = 0, \cdots, P_j(x_n) = 0$.

By the Lagrange Interpolation Formula, the jth row of A^{-1} is composed of the coefficients of the jth Lagrange basis polynomial:

$P_j(x) = \sum_{k=1}^{n} a_{kj} x^{k-1} = \prod_{\substack{1\le m\le n \\ m\neq j}} \frac{x - x_m}{x_j - x_m}$. In these two polynomials, it yields: $a_{kj} = (-1)^{n-k}\left(\sum_{\substack{1\le m_1<\cdots<m_{n-k}\le n \\ m_1,\cdots,m_{n-k}\neq j}} x_{m_1}\cdots x_{m_{n-k}}\right) \Big/ \prod_{\substack{1\le m\le n \\ m\neq j}}(x_j - x_m)$ $(1 \le k < n)$, $1 \Big/ \prod_{\substack{1\le m\le n \\ m\neq j}}(x_j - x_m)$ $(k = n)$.

And $w_i = \sum_{j=1}^{n} a_{ij} y_j$ $(i = 1, 2, \cdots, n)$

2. (20pt)

Suppose that $n-1 > d$. 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)

Since has only a trivial solution, matrix A has linearly independent columns. By the definition of Moore-Penrose inverse A^+ and state that A has linearly independent columns, thus matrix A^TA is invertible, $A^+ = (A^TA)^{-1}A^T$. So the equation $Aw = y$ is equivalent to:

$A^TAw = A^Ty$, $\underline{w = (A^TA)^{-1}A^Ty}$, $W = A^+y$ (A^+ is pseudoinverse of A).

$$c_1\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} + c_2\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} + c_3\begin{bmatrix}x_1^2\\ x_2^2\\ \vdots\\ x_n^2\end{bmatrix} + \cdots + c_{d+1}\begin{bmatrix}x_1^d\\ x_2^d\\ \vdots\\ x_n^d\end{bmatrix} = 0$$

that is, when $c_1 = c_2 = c_3 = \cdots = c_{d+1} = 0$, equation is only consistent,