Polynomial Regression (Handwriting Assignment)

Name: Héloïse ROBIN (헬로이스 로빈)

Student ID: 50231639

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 nth degree polynomial in x.

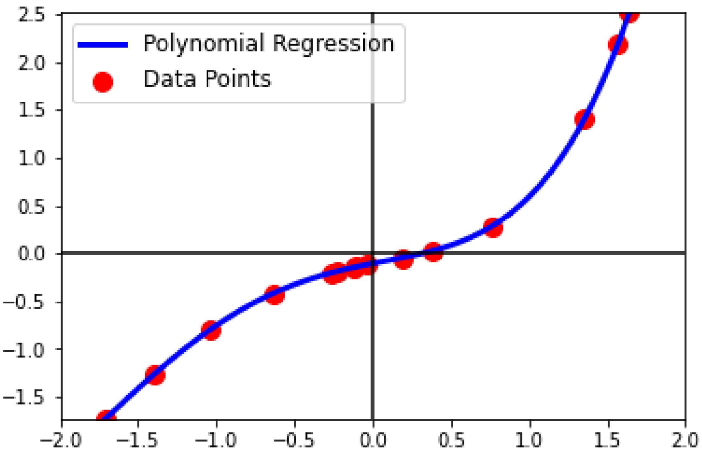


Figure 1: Example of Polynomial Regression

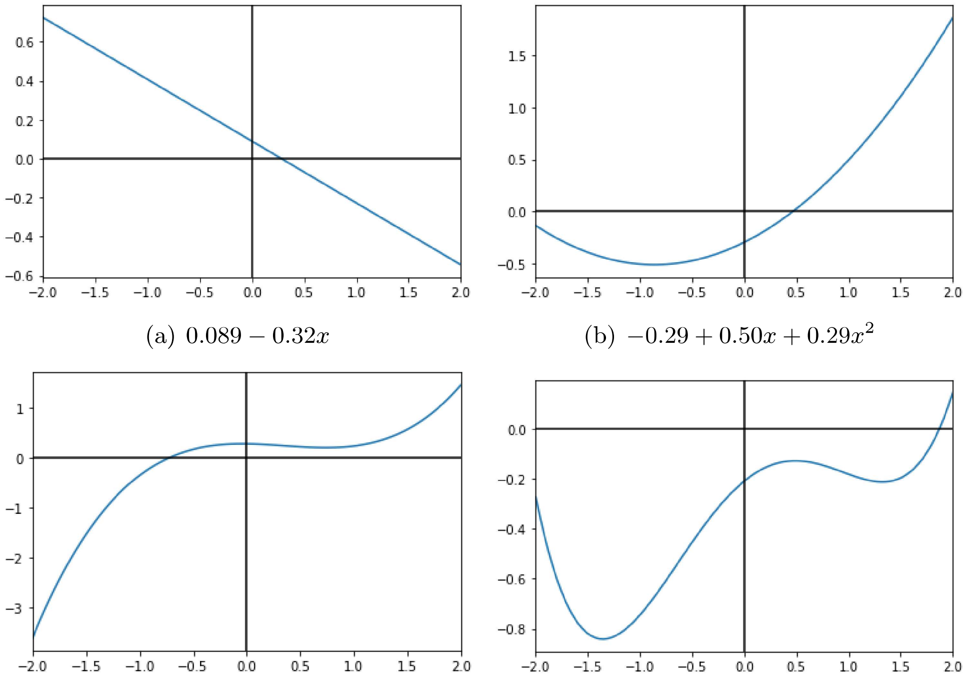
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

We have the equation system as followed :

$$\begin{cases} \hat{f}(x_1) = w_0 + w_1 x_1 + \dots + w_d x_1^d = y_1 \\ \hat{f}(x_2) = w_0 + w_1 x_2 + \dots + w_d x_2^d = y_2 \\ \vdots \\ \hat{f}(x_m) = w_0 + w_1 x_m + \dots + w_d x_m^d = y_m \end{cases}$$

This system can be written as a linear equation form such as $Aw = y$ with the matrix A and the vectors w and y.

A is the stack of $[1, x_i, x_i^2, \dots, x_i^d]$ for $i = 1, \dots, m$.

In this form, w is a vector representing all the coefficients w_j with $j = 0, \dots, d$, so $d+1$ elements.

$\Rightarrow$ $w = [w_0, \dots, w_d]^T$

$\Rightarrow$ w is a vector with $d+1$ lines and 1 column

On the same logic, y is a vector representing the elements y_i with $i = 1, \dots, m$, so m elements.

$\Rightarrow$ $y = [y_1, y_2, \dots, y_m]^T$

$\Rightarrow$ y is a vector composed of m lines and 1 column

$$\Rightarrow \begin{cases} \dim(w) = d+1 \\ \dim(y) = m \end{cases}$$

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

By the linear equation form, let A a matrix as such :

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ a_{n1} & a_{n2} & & \cdots & & a_{nd} \end{pmatrix}$$

Then, $Aw = y \Leftrightarrow \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ a_{21} & & \\ \vdots & \ddots & \\ a_{n1} & & a_{nd} \end{pmatrix} \begin{pmatrix} w_0 \\ \vdots \\ w_d \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

By posing the system and by identification, we can estimate the a_{ij} cofficients of the matrix A.

Moreover, A is said to be the stack of $[1, x_i, x_i^2, \dots, x_i^d]$ for $i = 1, \dots, m$.

Then, A is a matrix formed by m lines and $d+1$ columns.

$\Longrightarrow$ A is a matrix of size $(m, d+1)$

$$\Longrightarrow \quad A = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & \cdots & x_2^d \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 & \cdots & x_3^d \\ \vdots & & & \ddots & & & \vdots \\ 1 & x_m & x_m^2 & x_m^3 & x_m^4 & \cdots & x_m^d \end{pmatrix} \; m$$

$\underbrace{1}_{1} \; \underbrace{\quad\quad\quad}_{d}$

$\longleftrightarrow$ $d+1$

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

Let $d+1=n$

$\Rightarrow A(n, d+1) \Rightarrow A(n,n)$

Therefore, A becomes a square matrix and its explicit expression changes too.

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

Two methods are then possible: first with the Vandermonde matrix properties, and second with the "traditional" computation.

(1) Vandermonde Method

We can recognize a Vandermonde matrix pattern-like in the explicit expression of A. Indeed, as a Vandermonde matrix is a square matrix with the same number of rows and columns, and the entries of each row being the powers of x_i, we can affirm that A is a Vandermonde matrix when $d+1=n$. This statement comes to be really useful as we can now use some properties of this special kind of matrix to compute the determinant of A.

Indeed, for a Vandermonde matrix V, $\det(V)$ can be expressed as the product of the differences between the x_i values.

$$\Rightarrow \det(V) = \prod_{0 \leq i < j \leq n} (x_j - x_i)$$

We have directly the following:

$$\det(A) = (x_2-x_1)(x_3-x_1)(x_3-x_2)\ldots(x_n-x_1)\ldots(x_n-x_{n-1})$$

(2) Traditional "By hand" Method

$$\det(A) = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \ddots & & \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 0 & x_2-x_1 & x_2^2-x_1^2 & \cdots & x_2^{n-1}-x_1^{n-1} \\ \vdots & & \ddots & & \\ 0 & x_n-x_1 & x_n^2-x_1^2 & \cdots & x_n^{n-1}-x_n^{n-1} \end{vmatrix} \begin{array}{l} L_1 \\ L_2-L_1 \\ \vdots \\ L_n-L_1 \end{array}$$

$$= \prod_{2 \leq j \leq n} (x_j - x_1) \cdot \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & x_2^{n-} \\ \vdots & \vdots & & \\ 0 & 1 & \cdots & x_n^{n-} \end{vmatrix}$$

$\vdots$ (etc)

$$\det(A) = \prod_{0 \leq i < j \leq n} (x_j - x_i)$$

In conclusion, we find the same expression as in the Vandermonde matrix's determinant property.

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

We have $\det(A) = \prod_{0 \leq i < j \leq n} (x_j - x_i)$, a polynomial.

Then, we can write the following: if $\det(A) = 0$, then

$\prod_{0 \leq i < j \leq n} (x_j - x_i) = 0 \iff \exists (i,j) \in \mathbb{N} \,/\, 0 \leq i < j \leq n,\ x_j = x_i$

As such, we can now define the condition on which $\det(A) \neq 0$ (i.e. A invertible).

The determinant of A is non-zero IF AND ONLY IF all x_i are distinct (i.e. not equal).

$\det(A) \neq 0 \iff \forall (i,j) \in \mathbb{N} \,/\, 0 \leq i < j \leq n,\ x_j \neq x_i$

$\Rightarrow A$ invertible $(\exists A^{-1})$

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)

We saw in the precedent question that $\det(A) \neq 0$ for A a square Vandermonde matrix.

$\Rightarrow A$ invertible such as $\exists A^{-1} \mid A^{-1}A = AA^{-1} = I$

Then, to solve $Aw = y$, with respect to w, we can use A^{-1}, the inverse matrix of A as followed:

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

$\iff Iw = A^{-1}y$

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

With knowledge of A and y we can theoratically solve the equation. We now need to determine A^{-1}.

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A) = \frac{1}{\det(A)} (\operatorname{com}(A))^t$$

And we have:

$$\operatorname{adj}(A) = (-1)^{i+j} \det(A_i(x_1, \dots x_{j-1}, x_{j+1}, \dots x_m))$$

$$= (-1)^{i+j} S_{m-i,j} \prod_{\ell < k (\ell, k \neq j)}^{m} (x_k - x_\ell)$$

$$\Rightarrow A^{-1} = \frac{1}{\prod_{0 \leq i < j \leq n} (x_j - x_i)} \cdot (-1)^{i+j} S_{m-i,j} \prod_{\ell < k (\ell, k \neq j)}^{m} (x_k - x_\ell)$$

(I wasn't able to compute A^{-1} so I am stuck here ...)

Let a_{ij}^{-1} the coefficient of A^{-1} matrix.

We have: $A^{-1} = \begin{pmatrix} a_{11}^{-1} & a_{12}^{-1} & a_{13}^{-1} & \dots & a_{1n}^{-1} \\ a_{21}^{-1} & a_{22}^{-1} & a_{23}^{-1} & \dots & a_{2n}^{-1} \\ \vdots & & \ddots & & \vdots \\ a_{n1}^{-1} & & \dots & & a_{nn}^{-1} \end{pmatrix}$

$$\Rightarrow w = A^{-1}y \iff \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{n-1} \end{pmatrix} = \begin{pmatrix} a_{11}^{-1} & a_{12}^{-1} & \dots & a_{1n}^{-1} \\ a_{21}^{-1} & a_{22}^{-1} & \dots & a_{2n}^{-1} \\ \vdots & & \ddots & \vdots \\ a_{n1}^{-1} & & \dots & a_{nn}^{-1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\iff \begin{cases} w_0 = a_{11}^{-1} y_1 + a_{12}^{-1} y_2 + \dots + a_{1n}^{-1} y_n \\ w_1 = a_{21}^{-1} y_1 + a_{22}^{-1} y_2 + \dots + a_{2n}^{-1} y_n \\ \vdots \\ w_n = a_{n1}^{-1} y_1 + a_{n2}^{-1} y_2 + \dots + a_{nn}^{-1} y_n \end{cases}$$

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

When the matrix A is not square (i.e., when $n>d+1$), we can't compute A^{-1} directly, which makes solving the linear equation $Aw=y$ using the matrix inverse method impractical.

As $n>d+1$, we have more equations than unknowns, so we want to deal with an overdetermined systems.

However, we can still solve this overdetermined system of linear equations by using various techniques, such as the method of least squares

We can use the approach as followed:

(1) The least squares solution finds a vector w that minimizes the sum of squared differences between the product Aw and the vector y.

To calculate w we need to

- compute A^t
- compute $A^t A$
- inverse the precedent result to get $(A^tA)^{-1}$

So we then can use the formula:
$w = (A^tA)^{-1} A^t y$

(2) Another way to approach the problem is by using the pseudoinverse of A to find the least squares solution. The pseudoinverse is a tool that accomodates non-square matrices:
$w = A^+ y$

These approaches are widely employed when dealing with overdetermined systems, where you have more equations than unknowns. These methods can be implemented using numerical linear algebra libraries available in programming languages like Python (NumPy), MATLAB, or other mathematical software.