

Polynomial Regression (Handwriting Assignment)

Name: 이승재

Student ID: 20236004

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 n th degree polynomial in x .

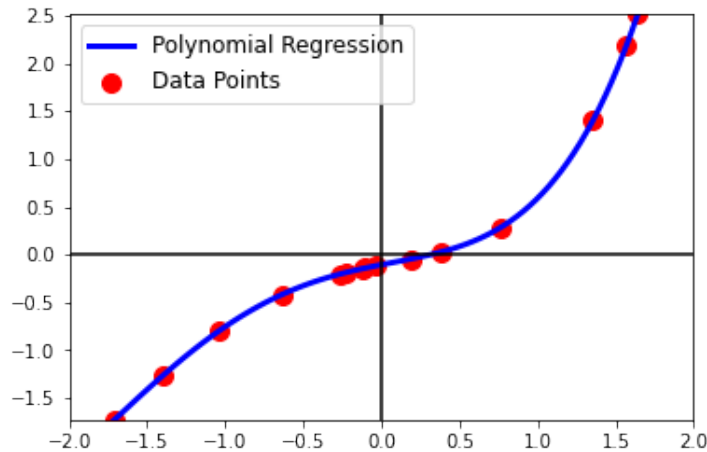


Figure 1: Example of Polynomial Regression

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,

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

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

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

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:

$$\text{Degree of 0 : } f(x) = w_0$$

$$\text{Degree of 1 : } f(x) = w_1 \cdot x + w_0$$

$$\text{Degree of 2 : } f(x) = w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\text{Degree of 3 : } f(x) = w_3 \cdot x^3 + w_2 \cdot x^2 + w_1 \cdot x + w_0$$

$$\vdots$$

$$\text{Degree of } d : f(x) = \sum_{i=0}^d w_i \cdot x^i,$$

where w_0, w_1, \dots, 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 $\{w_0, w_1, \dots, w_d\}$. 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, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Let the degree of polynomial be d . Then, we want to find $w_0, w_1, w_2, \dots, w_d$ of the polynomial such that

$$\begin{aligned}\hat{f}(x_1) &= w_0 + w_1x_1 + w_2x_1^2 + \dots + w_dx_1^d = y_1, \\ \hat{f}(x_2) &= w_0 + w_1x_2 + w_2x_2^2 + \dots + w_dx_2^d = y_2, \\ \hat{f}(x_3) &= w_0 + w_1x_3 + w_2x_3^2 + \dots + w_dx_3^d = y_3, \\ \hat{f}(x_4) &= w_0 + w_1x_4 + w_2x_4^2 + \dots + w_dx_4^d = y_4, \\ \hat{f}(x_5) &= w_0 + w_1x_5 + w_2x_5^2 + \dots + w_dx_5^d = y_5, \\ &\vdots \\ \hat{f}(x_n) &= w_0 + w_1x_n + w_2x_n^2 + \dots + w_dx_n^d = y_n.\end{aligned}$$

Now, we reformulate the equations into the vector and matrix form. First, let $\mathbf{w} = [w_0, w_1, \dots, w_d]^T$ and $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$. Then, the above equations can be rewritten as

$$\hat{f}(x_1) = [1, x_1, x_1^2, x_1^3, \dots, x_1^d] \cdot \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_d \end{bmatrix} = [1, x_1, x_1^2, x_1^3, \dots, x_1^d] \mathbf{w} = y_1$$

Similarly, we have,

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

Then, all equations can be written as the form of linear equation,

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

where A is the stack of $[1, x_i, x_i^2, x_i^3, \dots, x_i^d]$ for $i = 1, \dots, n$. Under this setting, answer the following questions.

1-(a) What is the size of vector w and y ? (10pt)

\vec{w} has $(d+1)$ weights as elements in one column, so the size of vector \vec{w} is $d+1$

\vec{y} has n outputs as elements in one column, so the size of vector \vec{y} is n .

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

Matrix A has n rows depending on the number of outputs.

And it has $(d+1)$ columns depending on degree of polynomial function \hat{f} .

Therefore, the size of matrix A is $n(d+1)$.

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

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}$$

To find its determinant, use cofactor expansion because A is square matrix.

$$|A| = \sum_{j=1}^n (-1)^{i+j} a_{ij} |M_{ij}|,$$

M_{ij} is a submatrix of A , which consists of the remaining elements except for the i -th row and j -th column in A

Since M_{ij} is also a square matrix, its determinant can be obtained by using cofactor expansion. By repeating this process recursively, the determinant of A can be obtained.

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

According to theorem of invertible matrix, A does not have 0 as its eigenvalue, and the column vectors of A are linearly independent.

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

Since A is an invertible matrix, the inverse of A can be obtained.

$$A^{-1}A\vec{w} = A^{-1}\vec{y}, \quad A^{-1}A = I, \quad \text{thus } \vec{w} = A^{-1}\vec{y}.$$

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

Find a pseudoinverse matrix of A , A' .

A' satisfies $A'A = I_{d+1}$, $AA' = I_n$.

To find A' , do singular value decomposition (SVD) for A .

Assume $A = U\Sigma V^T$, U and V are orthogonal matrix, Σ is diagonal matrix including singular values of A .

A is tall matrix ($n > d+1$), assume $\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \sigma_{d+1} \\ \vdots & \vdots & & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix}$, $\Sigma' = \begin{bmatrix} \sigma_1^{-1} & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & 0 & & \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & & \sigma_{d+1}^{-1} & 0 \dots 0 \end{bmatrix}$ n d+1

Then $\Sigma'\Sigma = I_{d+1}$.

U and V are orthogonal, $U^T A = U^T U \Sigma V^T = \Sigma V^T$, $\Sigma' U^T A = \Sigma' \Sigma V^T = V^T$, $V \Sigma' U^T A = I_{d+1}$.

$\therefore A' = V \Sigma' U^T$

Thus, $A\vec{w} = \vec{y}$, $A'A\vec{w} = I\vec{w} = \vec{w} = A'\vec{y}$.