Polynomial Regression (Handwriting Assignment)

Name: Valentin Durieux

Student ID: 50231621

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.

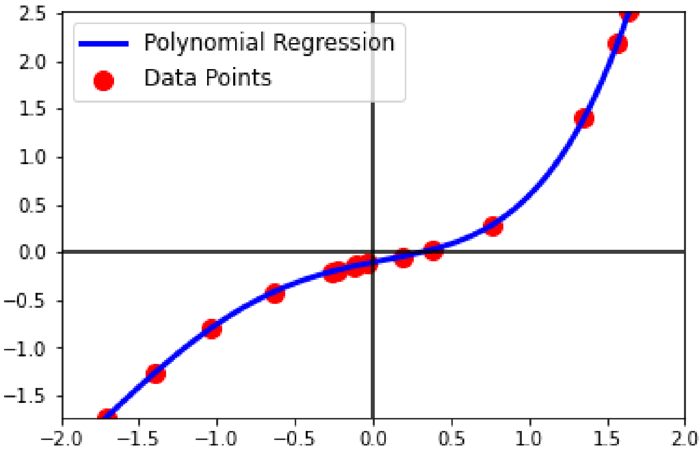


Figure 1: Example of Polynomial Regression

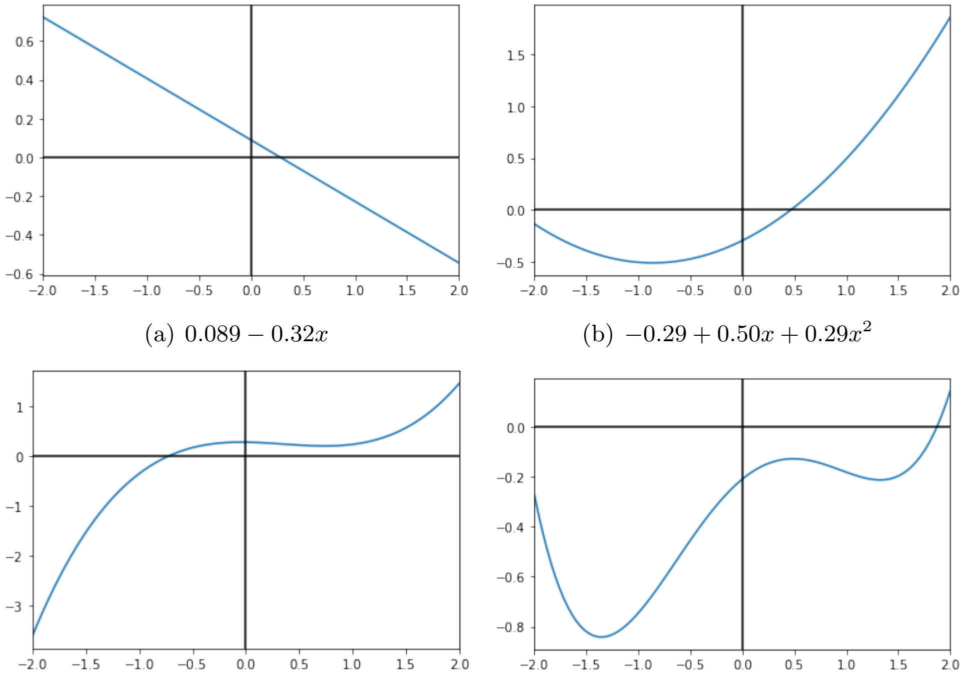
First, what is a regression? we can find a definition from the book as follows: *Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.* Actually, this definition is a bookish definition, in simple terms the regression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

Size of vector w:

Vector w contains the coefficients of the polynomial, and you want to find w0, w1, w2, ..., wd. Since the degree of the polynomial is d, w will have (d + 1) elements, where:
Size of w = d + 1

Size of vector y:

Vector y contains the target values for each data point. You have n data points, so vector y will have n elements, one for each data point.
Size of y = n

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

The matrix A is formed by stacking the vectors [1; xi; x2i; x3i; ...; xdi] for each data point i = 1, 2, ..., n. The size of matrix A will be n x (d + 1), where n is the number of data points, and (d + 1) is the size of vector w.

So, the size of matrix A is n x (d + 1).

To write out the matrix A explicitly, it would look like this:

```
A =
| 1   x1   x1^2   x1^3   ...   x1^d   |
| 1   x2   x2^2   x2^3   ...   x2^d   |
| 1   x3   x3^2   x3^3   ...   x3^d   |
| 1   x4   x4^2   x4^3   ...   x4^d   |
| 1   x5   x5^2   x5^3   ...   x5^d   |
| ...                          |
| 1   xn   xn^2   xn^3   ...   xn^d   |
```

Each row of this matrix corresponds to one data point, and each column corresponds to a different term in the polynomial of degree d.

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

det(A) = $\prod(x_i - x_j)$, for all $i \neq j$ from 1 to n.

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

The determinant of matrix A is a product of differences between the data points, and it will be nonzero if and only if the data points $x1, x2, \ldots, xn$ are all distinct.

In other words, for the determinant of A to be non-zero, no two data points can be the same; each xi must be different from all other xj for $i \neq j$.

If any of the data points are repeated or not distinct, the determinant of A will be zero.

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)

If the determinant of matrix A is non-zero, that means the matrix A is invertible.
In this case, we can solve the linear equation $Aw = y$ for w by taking the inverse of A:

$w = A^{-1} * y$

So, if the determinant of A is non-zero, the solution for w with respect to the linear equation $Aw = y$ is given by $w = A^{-1} * y$, where A^{-1} is the inverse of the matrix A.

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

If n > d + 1, and A is not a square matrix, we can't directly compute the inverse of A to solve the linear equation Aw = y.
In this scenario, we can use techniques such as the least squares method to find an approximate solution to the equation.

One common approach is to use the pseudoinverse of A, denoted as A^+. The solution for w can be found as follows:

$w = A^+ * y$

The pseudoinverse A^+ is a generalization of the matrix inverse for non-square matrices and is used to find the "best fit" solution to the overdetermined system of linear equations. This approach minimizes the least squares error between the predicted values Aw and the actual values y.

In practice, numerical computing tools and libraries often provide functions for solving linear equations using the least squares method, which makes it easier to implement without explicitly calculating the pseudoinverse.