Polynomial Regression

Regression task

In regression task, the problem is to approximate an unknown real valued target function, such that:

$$y = f(x) + \varepsilon$$

y: output data.

x: input data.

f: real valued target function.

ε is a centered white noise with $E[\varepsilon] = 0$ and $Var(\varepsilon) = \sigma^2$.

Linear Model

Definition: linear model

A model is said to be linear if it is linear in parameters. Linear model is featured by the following hypothesis:

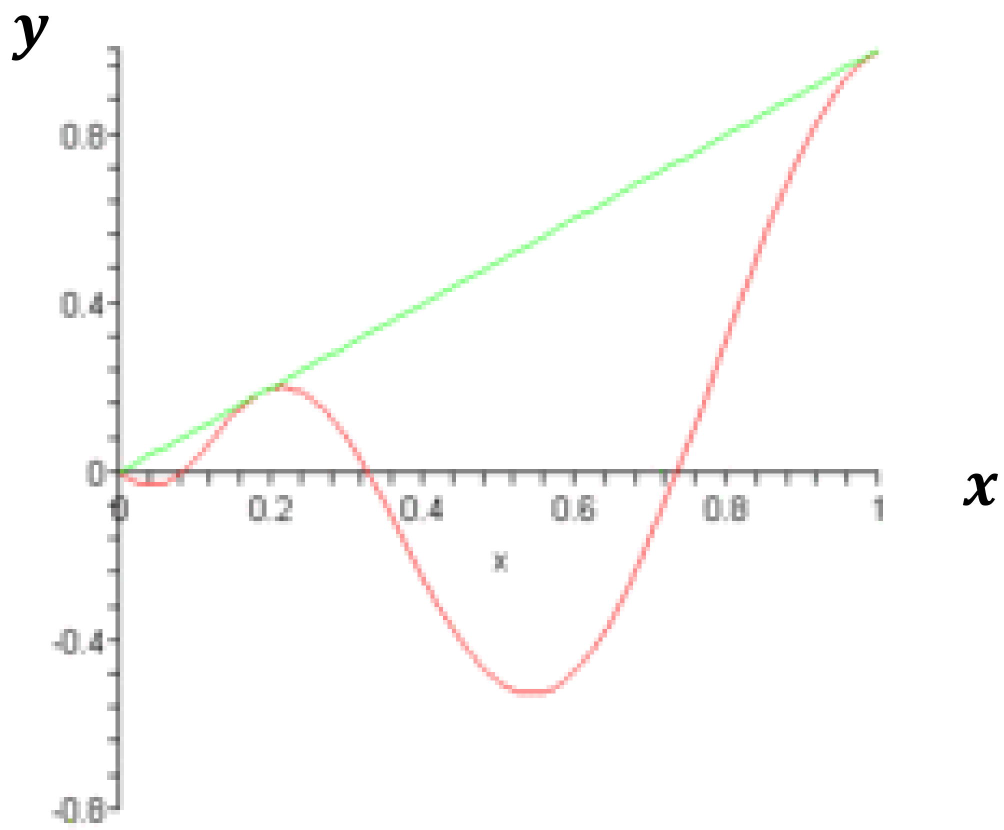
$$h_w(x) = \sum_{i=0}^{d} w_i x_i$$

Linear model is characterized by a linear class of hypothesis.

Motivation

Objective:

Explain the quantitative variable y by the d variables $x = (x_1, \ldots, x_d)$.



Tool:

Polynomial Regression

Polynomial Regression

To approximate f:

- We should define a real-valued hypothesis class H based on prior knowledge.
- We should find the best hypothesis that has small general risk.

Let’s consider that the prior knowledge assumes that the relationship between the outputs y and the inputs x is polynomial.

Polynomial Regression

The hypothesis class for polynomial regression model:

Therefore, H will be defined as a class of polynomial hypotheses.

$$H = \{h_W : X \to \mathbb{R} \ \ , W \in \mathbb{R}^{D+1}\}$$

$$h_W(x) = w_0 + w_1 x_1 + \cdots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \cdots + w_r x_d^2 +$$

$$\ldots + w_s x_1^Q + w_{s+1} x_1^{Q-1} x_2 + \cdots + w_D x_d^Q$$

This is a multi-dimensional polynomial regression of order Q.

Polynomial Regression

Indeed, h_w appears to be nonlinear, however, it is linear in parameters W.

Let's take:

$$z_1 = 1$$
$$z_2 = x_1$$
$$\vdots$$
$$z_d = x_d$$
$$z_{d+1} = x_1^2$$
$$z_{d+2} = x_1 x_2$$
$$\vdots$$

$$z_r = x_d^2$$
$$\vdots$$
$$z_s = x_1^Q$$
$$z_{s+1} = x_1^{Q+1} x_2$$
$$\vdots$$
$$z_D = x_d^Q$$

Polynomial Regression

So, the hypothesis becomes:

$$h_W(z) = w_0 z_0 + w_1 z_1 + \cdots + w_d z_d + \cdots + w_r z_r + \cdots + w_s z_s + \cdots + w_D z_D$$

Which is linear in parameters W.

Definition: Polynomial regression model

It is a linear model used to capture curvature in data by using higher-order values of inputs. It is a linear combination of higher-ordered inputs:

$$h_W(x) = w_0 + w_1 x_1 + \cdots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \cdots + w_D x_d^Q$$

It is also called **curvilinear** regression.

Parameter Estimation

To estimate the parameters of polynomial regression, we should minimize the following loss function:

$$L_S(w) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_w(x^{(i)})\right)^2$$

Which implies:

$$L_S(w) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \sum_{j=0}^{D} w_j z_j^{(i)}\right)^2$$

Parameter Estimation

So, the loss function becomes:

$$L_S(w) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - w^T . z^{(i)}\right)^2$$

To optimize this function, we can use an iterative algorithm named Gradient Descent:

$$\nabla L_S\left(w_j\right) = \frac{-2}{m}\sum_{i=1}^{m}\left(y^{(i)} - w^T . z^{(i)}\right) z_j^{(i)}$$

$$w_j \leftarrow w_j - \alpha \nabla L_S\left(w_j\right)$$

Gradient Descent Algorithm for Polynomial Regression

Input: The training data: $S = \{\left(x^{(1)}, y^{(1)}\right), \dots, (x^{(m)}, y^{(m)})\}$

The polynomial order Q, The estimation parameter ε

The initial vector of parameters $w^{(0)}$ and the learning rate α

Output: w^*, t and $L_S(w^*)$

Start: $w \leftarrow w^{(0)}$

Compute the value of the hypothesis for each observation x :

$$h_W(x) = w_0 + w_1 x_1 + \cdots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \cdots + w_D x_d^Q$$

$$= \sum_{j=0}^{D} w_j z_j$$

Compute the cost function: $L_S(w^{(0)}) = \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - (w^{(0)})^T . z^{(i)}\right)^2$

$t = 0$

While $(L_S\left(w^{(t)}\right) > \varepsilon)$ {

$$\boldsymbol{w^{(t+1)}} \leftarrow \boldsymbol{w^{(t)}} - \boldsymbol{\alpha \nabla L_S\left(w^{(t)}\right)}$$

$t \leftarrow t + 1$ }

$$w^* \leftarrow w^{(t)}$$

Return: w*, t and $L_S(w^*)$

End