

Polynomial Regression

Regression task

In regression task, the problem is to approximate an unknown real valued target function, $(x_i, y_i), i \in I$ such that: we are looking for $h \in H$

With probability we have $f(x_i) = \mathbf{y_i(\varepsilon_i)} = \mathbf{h_S(x_i)} + \mathbf{\varepsilon_i}$

- $cov(\varepsilon_i, \varepsilon_{i+1}) = 0$, $\{\varepsilon_i\}$ are independent and identically distributed (**iid**)
- $S = \{(x_i, y_i), i \in I\}$ training set $|I| = m$
- $h_S = \operatorname{argmin}_{h \in H} L_S(h) \Rightarrow \text{if } L_S(h_S) = 0 \Leftrightarrow \forall i \in I \text{ } \mathbf{h_S(x_i)} = y_i$
- $\mathbf{y_i(\varepsilon_i)} = y_i \in \mathbb{R}$: output data, Label of $x \in \mathbb{R}^d$
- $x \in \mathbb{R}^d$: input data.
- f : stochastic real valued target function.
- ε_i is a centered white noise with $E[\varepsilon_i] = 0, cov(\varepsilon_i, \varepsilon_{i+1}) = 0$, and $Var(\varepsilon_i) = \sigma^2$.

Regression task

- Input :
 - $S = \{(x_i, y_i), i \in I\}$ training set $|I| = m$, $x_i = (x_i^1, x_i^2) \in \mathbb{R}^2$ and $y_i \in \mathbb{R}$
 - $L_S(h) = L_S(w)$, $w \in \mathbb{R}^{D(Q)}$

$$H = \{h_{D,Q}\} = \{w \in \mathbb{R}^D : h_{D,Q}\} \Rightarrow |H| \approx \infty,$$

$$h_{D,Q}(x) = w_0 + w_1 x_1 + \dots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \dots + w_r x_d^2 + \dots + w_s x_1^Q + w_{s+1} x_1^{Q-1} x_2 + \dots + w_D x_d^Q$$

- $x = (x_1, \dots, x_d) \in \mathbb{R}^d$

- Output

$$h_S = \operatorname{argmin}_{h \in H} L_S(h) \Leftrightarrow w^* = \operatorname{argmin}_{w \in \mathbb{R}^Q} L_S(w)$$

- $h_{4,2}(x_i) = w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^1)^2 + w_4 (x_i^2)^2 \rightarrow \nabla h_{4,2}(x_i) = (x_i^1, x_i^2, x_i^1 x_i^2, (x_i^1)^2, (x_i^2)^2)^T$
- $h_{3,2}(x_i) = w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^1)^2$
- $h_{3,2}(x_i) = w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^2)^2$
- $h_{2,2}(x_i) = w_0 x_i^1 + w_1 x_i^2 + w_2 (x_i^2)^2$

d=2

- $\mathbf{h}_{4,2}(x_i) = \text{Bais} + w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^1)^2 + w_4 (x_i^2)^2$
 $\rightarrow \nabla \mathbf{h}_{4,2}(x_i) = (1, x_i^1, x_i^2, x_i^1 x_i^2, (x_i^1)^2, (x_i^2)^2)^T$
- $\mathbf{h}_{3,2}(x_i) = \text{Bais} + w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^1)^2$
- $\mathbf{h}_{3,2}(x_i) = \text{Bais} + w_0 x_i^1 + w_1 x_i^2 + w_2 x_i^1 x_i^2 + w_3 (x_i^2)^2$
- $\mathbf{h}_{2,2}(x_i) = \text{Bais} + w_0 x_i^1 + w_1 x_i^2 + w_2 (x_i^2)^2$

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: $(x, y) = (x_j, y_j)$

For each j we have $x = x_j = (x_1, \dots, x_d)$

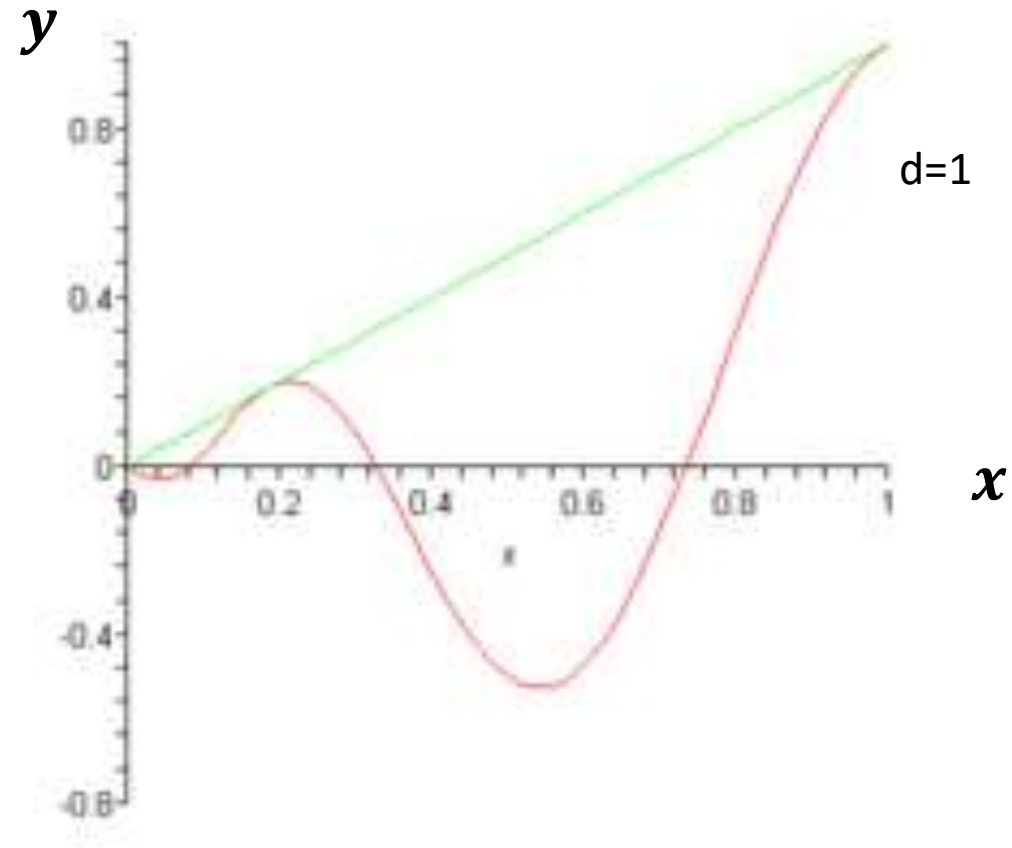
$$h(x) = \sum_{i=0}^d w_i x_i$$

Linear model is characterized by a linear class of hypothesis.

Motivation

Objective:

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



Tool:

Polynomial Regression

Polynomial Regression: A_α

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 A_α

The hypothesis class for polynomial regression model:

Input: $d, x_i = x = (x_0, x_1, \dots, x_d) \in \mathbb{R}^{d+1}, x_0 = 1$

Hyper Parameters: $\alpha = (D(Q), Q)$, Aim: finding the target function that means finding the best hyper parameters D, Q

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

$$H^{D(Q),Q} = \{h_{D(Q),Q}: X \rightarrow \mathbb{R} \text{ , } \mathbf{w} \in \mathbb{R}^{D(Q)+1}\} \Rightarrow |H^{D,Q}| = \infty$$

$$h_{D(Q),Q}(x) = w_0 + w_1x_1 + \dots + w_dx_d + w_{d+1}x_1^2 + w_{d+2}x_1x_2 + \dots + w_rx_d^2 + \\ \dots + w_sx_1^Q + w_{s+1}x_1^{Q-1}x_2 + \dots + w_Dx_d^Q$$

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

$$\mathbf{x} = (x_0, x_1, \dots, x_d)$$

$$h_{D(Q),Q}(x) = w_0 + P_{D(Q),Q}(x)$$

- $d = 3 \Rightarrow x = (1, x_1, x_2, x_3)$
- $Q=2, D=9$
- $P_{9,2}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1x_2 + w_5x_1x_3 + w_6x_2x_3 + w_7x_1^2 + w_8x_2^2 + w_9x_3^2$
- $Q=2, D=7$
 - $P_{2,7}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1x_2 + w_5x_1x_3 + w_6x_2x_3 + w_7x_1^2$
 - $P_{2,7}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_2x_3 + w_5x_1^2 + w_6x_2^2 + w_7x_3^2$
- $Q=2, D=5$

$$h_{D(Q),Q}(\mathbf{x}) = \mathbf{w}_0 + \mathbf{P}_{D(Q),Q}(\mathbf{x})$$

$$h_{D(Q),Q}(\mathbf{x}) = \mathbf{w}_0 + \mathbf{w}^T \psi_Q(x_1, \dots, x_d) \quad \mathbf{w} \in \mathbb{R}^D$$

$$\psi_Q(x_1, \dots, x_d) = \begin{pmatrix} x_1, \dots, x_d \\ x_1^2, \dots, x_d^2 \\ \dots \\ x_1^Q, \dots, x_d^Q \end{pmatrix}$$

$$\mathbf{x} = (x_1, \dots, x_d) \rightsquigarrow \mathbf{z} = (x_1^2, x_1x_2, \dots, x_1x_d, x_2^2, x_2x_3, \dots, x_2x_d, x_3^2, \dots, x_{d-1}^2, x_{d-1}x_d, \dots, x_d^2)$$

Polynomial Regression

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

Let's take:

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

$$\begin{aligned} z_r &= x_d^2 \\ &\vdots \\ z_s &= x_1^Q \\ z_{s+1} &= x_1^{Q-1} x_2 \\ &\vdots \\ z_D &= x_d^Q \end{aligned}$$

Polynomial Regression

So, the hypothesis becomes:

$$h_{D(Q)}(z) = w_0 z_0 + w_1 z_1 + \cdots + w_d z_d + \cdots + w_r z_r + \cdots + w_s z_s + \cdots + w_D z_D$$

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:

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

It is also called **curvilinear** regression.

Parameter Estimation

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

$$L_S(h_{D(Q),Q}) = \frac{1}{m} \sum_{i=1}^m (y^{(i)} - h_{D(Q),Q}(x^{(i)}))^2$$

Which implies:

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

Parameter Estimation

So, the loss function becomes:

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

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

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

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

Gradient Descent Algorithm for Polynomial Regression

- **Input:**

1. The training data: $S = \{(x^{(0)} = 1, y^{(0)} = w_0), (x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$
2. The polynomial order Q , **D(Q): dimension of weight**, the Learning rate α , The estimation parameter ϵ
3. The initial vector of parameters $w^{(0)}$ and the learning rate α

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

- **Start:**

1. $w \leftarrow w^{(0)}$
2. Compute the value of the hypothesis for each observation x :

$$h_{D(Q)}(x) = w_0 + w_1x_1 + \dots + w_dx_d + w_{d+1}x_1^2 + w_{d+2}x_1x_2 + \dots + w_Dx_d^Q = \sum_{j=0}^D w_jz_j$$

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

- **While** (**$L_S(w^{(t)}) > \epsilon$**) {

1. Calculate α and $\nabla L_S(w^{(t)})$
2. $w^{(t+1)} \leftarrow w^{(t)} - \alpha \nabla L_S(w^{(t)})$
3. $t \leftarrow t + 1$ }
4. $w^* \leftarrow w^{(t)}$

- **Return:** w^* , t and $L_S(w^*)$, D^* : dimension of weight $w^* = (w_1, w_2, \dots, w_D)$ that $w_i \neq 0$ if all $w_i \neq 0$ then $D^* = D$

End