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 $f(x_i) = \boldsymbol{y_i(\varepsilon_i) = h_S(x_i) + \varepsilon_i}$

- $cov(\varepsilon_i,\, \varepsilon_{i+1}) = 0$, $\{\varepsilon_i\}$ are independent and identically distributed (**iid**)
- $S = \{(x_i, y_i), i \in I\}$ trainning set $|I| = m$
- $h_S = \underset{h \in H}{\operatorname{argmin}} L_S(h) \Longrightarrow if\ L_S(h_S) = 0 \Leftrightarrow \forall i \in I\ \ h_S(x_i) = y_i$
- $y_i(\varepsilon_i) = y_i \in \mathbb{R}$: 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 :
  - $S = \{(x_i, y_i), i \in I\}$ trainning set $|I| = m$, $x_i = (x_i^1, x_i^2) \in \mathbb{R}^2$ and $y_i \in \mathbb{R}$
  - $L_S(h) = L_S(w)$, $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{D(Q)}}$

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

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

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

- Output

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

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

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 $x = x_j = (x_1, \cdots, x_d)$

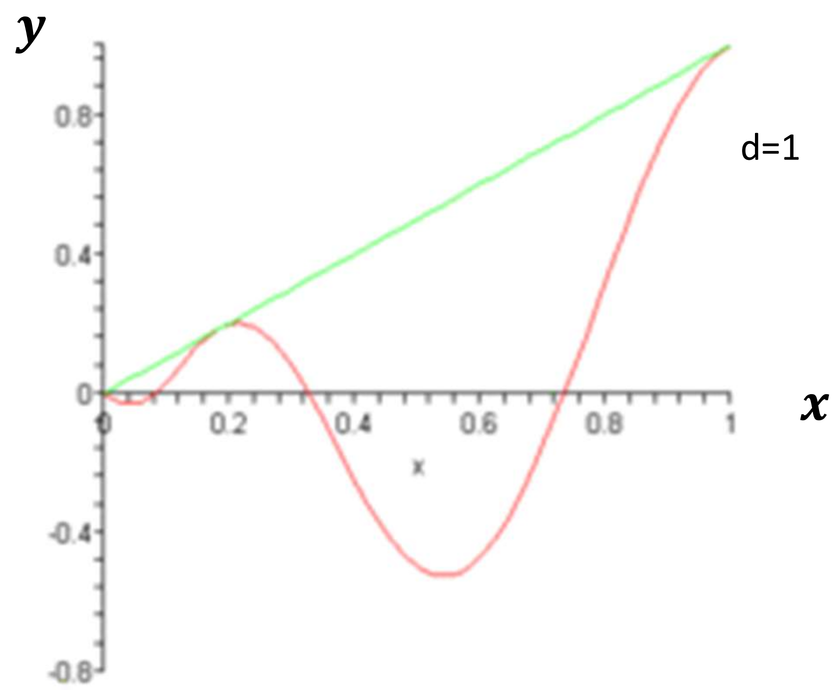
$$h(x) = \sum_{i=0}^{d} w_i x_i$$

Linear model is characterized by a linear class of hypothesis.

Motivation

Objective:

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



Tool:

Polynomial Regression

Polynomial Regression: A_α

To approximate $\boldsymbol{f}$:

- We should define a real-valued hypothesis class $\boldsymbol{H}$ 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: $\alpha = (\boldsymbol{D(Q), Q})$, Aim: finding the target function that means finding the bests hyper parameters D, Q

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

$$H^{D(Q),Q} = \{h_{D(Q),Q}: X \to \mathbb{R} \ \ , \boldsymbol{w} \in \mathbb{R}^{\boldsymbol{D(Q)+1}}\} \Longrightarrow |H^{D,Q}| = \infty$$

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

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

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

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

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

- $d = 3 \Longrightarrow x = (1, x_1, x_2, x_3)$
- Q=2, D=9
- $P_{9,2}(x) = \; w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1x_2 + w_5x_1x_3 + w_6x_2x_3 + w_7{x_1}^2 + w_8{x_2}^2 + w_9{x_3}^2$

- Q=2, D=7
  - $P_{2,7}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1x_2 + w_5x_1x_3 + w_6x_2x_3 + w_7{x_1}^2$
  - $P_{2,7}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_2x_3 + w_5{x_1}^2 + w_6{x_2}^2 + w_7{x_3}^2$

- Q=2, D=5

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

$$h_{D(Q),Q}(\boldsymbol{x}) = \boldsymbol{w_0} + \boldsymbol{w^T}\psi_Q(x_1, \dots, x_d) \;\; w \in \mathbb{R}^D$$

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

$$x_1^2, \dots, x_d^2 \rightsquigarrow x_1^2, x_1x_2, .., x_1x_d, x_2^2, x_2x_3, \dots, x_2x_d, x_3^2, \dots, x_{d-1}^2, x_{d-1}x_d, , x_d^2$$

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:

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

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:

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

It is also called **curvilinear** regression.

Parameter Estimation

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

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

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:

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

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

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

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

Gradient Descent Algorithm for Polynomial Regression

- **Input:**
  1. The training data: $S = \{(x^{(0)} = 1, y^{(0)} = w_0), (x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$
  2. The polynomial order Q, **D(Q): dimension of weight** , the Learning rate $\boldsymbol{\alpha}$, 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 :

$$h_{\mathrm{D(Q)}}(x) = w_0 + w_1 x_1 + \cdots + w_d x_d + w_{d+1} x_1^2 + w_{d+2} x_1 x_2 + \cdots + w_D x_d^Q = \sum_{j=0}^{D} w_j z_j$$

  1. Compute the cost function: $L_S(w^{(0)}) = \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - (w^{(0)})^T . z^{(i)}\right)^2$
  2. $t = 0$
- **While** ($\boldsymbol{L_S\left(w^{(t)}\right) > \varepsilon}$) {
  1. *Calculate* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\nabla L_S\left(w^{(t)}\right)}$
  2. $\boldsymbol{w^{(t+1)} \leftarrow w^{(t)} - \alpha \nabla L_S\left(w^{(t)}\right)}$
  3. $t \leftarrow t + 1$ }
  4. $w^* \leftarrow w^{(t)}$
- **Return:** w^*, t and $L_S(w^*)$, D^*: dimension of weight $\mathrm{w}^* = (w_1, w_2, \ldots, w_D)$ that $w_i \neq 0$ if all $w_i \neq 0$ then $D^* = \mathrm{D}$

End