Agenda :-
- Intro
- Linear Models
- Regression Solution
- Gradient Solution
- Loss functions
- Optimization
- Feature Engineering, overfit

Quick recap

Goal :
$y = f(x)$
$x \to$ data vector R^d
$y \to$ $\{0,1\}$
$f \to$ [ ]

$x^{(i)} = [x^{(i)}_1, x^{(i)}_2, x^{(i)}_3, \dots, x^{(i)}_d]$
data point

$f(x)$

Goal is also Get:

$y^{(i)} \to \boxed{\frac{f(x_i)}{model}} \to$ Predicts the output for data point $x^{(i)}$

Ground truth (actual value)

$y^{(i)} \to y^{(i)} = \frac{y^{(i)}}{\text{actual value}}$

$\hat{y}^{(i)} \to$ predicted output (hat)

$\hat{y}^{(i)} \to$ predicted value (hat)

Two phase

① Training phase

Using $\left\{\begin{matrix} \text{Historical data} \\ \text{Data set} \end{matrix}\right\}$ we find/fit the function $\boxed{f(x)}$

② Testing phase

Use $\boxed{\text{unseen data}}$ $\xrightarrow{\text{Input to}}$ $\boxed{\text{Model fit}}$ $\longrightarrow$ Predict the output

Intuition

Goal is to find weights

objective : $\boxed{f(x)}$ so $\hat{y} \approx y$

$\hat{y} \to$ Linear function

$f(x^{(i)}) \approx f(\begin{matrix} x^{(i)}_1, & x^{(i)}_2, & \dots & x^{(i)}_d \end{matrix})$

$\hat{y}^{(i)} = f(x^{(i)}) \approx w_1 x^{(i)}_1 + w_2 x^{(i)}_2 + w_3 x^{(i)}_3 + \dots + w_d x^{(i)}_d + w_0$

$\hat{y}^{(i)} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_d \end{bmatrix} \begin{bmatrix} x^{(i)}_1 \\ x^{(i)}_2 \\ x^{(i)}_3 \\ \vdots \\ x^{(i)}_d \end{bmatrix} + w_0$

$\hat{y}^{(i)} \quad \hat{y}(x) \approx w_1 x_1 + w_2 x_2 + w_0$

$\hat{y}(x) \approx w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_0$

$\hat{y}(x) \approx \boxed{2x_1 + 3x_2 + x_3 + \dots + 5x_4 + 3}$

$\hat{y} = f(x) \approx w_1 x_1 + w_2 x_2 + \dots + w_d x_d + w_0$

$\approx f(x, w_1, w_2, w_3, \dots, w_d) + w_0$

$\hat{y} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix} + w_0$

$\hat{y} = \boxed{w^T x + w_0}$

$\hat{y} \to$ Dot product

In order to find $f(x)$: find $w_1, w_2, w_3, \dots, w_d$ & w_0

Task : find best $\boxed{w, b}$

$f(x_{new}) = 2 \to w_1 x_{new}$
$2 \, x_{new} + 3 \to w_1 x_{new} + w_0$

$\to$ $w_1 = 2$ best

Linear Regression

① $\hat{y} = w_1 x + w_0$ $\to$ Fit line

$\hat{y}, f(x)$

$\hat{y} = w_1 x + w_0$ $\leftarrow$ $\hat{y} = f(x)$ a line

y

$f(x_{new}) \to 2.5 \, x_{new}$

② $\hat{y} = f(x) = w_1 x + w_0 + w_2 x^2$ $\to$ Fit curve

$\hat{y} = f(x) = w_1 x_1 + w_2 x_2 + w_0$

$\hat{y} = f(x)$ Plane.

$f(x_{new}) \to (4 \times 2) + 3.5 \, x_{new} + w_0$
approach the output

Non-linear Linear Regression

$y^{(i)} \to y^{(i)} = e^{x^{(i)}}$

$\hat{y}^{(i)} \to$ polynomial

$\hat{y}^{(i)} = \sum_{w_i, w_0} \left(\begin{matrix} \approx \\ = \end{matrix} \right) y^{(i)}$

find new w_i, w_0 (which minimizes the error)

$f(x_{new}) \to 2.5 \, x_{new}$

$\hat{y}^{(i)} = \sum_{w_i, w_0} \frac{1}{2} \sum_{i=1}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

$\hat{y}^{(i)} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_d \end{bmatrix} \begin{bmatrix} x^{(i)}_1 \\ x^{(i)}_2 \\ x^{(i)}_3 \\ \vdots \\ x^{(i)}_d \end{bmatrix} + w_0$

$\hat{y}^{(i)} = w^{(i)} = 2$

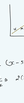
$\hat{y}^{(i)} = 0$

$\text{Error} = \frac{1}{2} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left| y^{(i)} - \hat{y}^{(i)} \right|$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss



$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left| y^{(i)} - \hat{y}^{(i)} \right|$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$x^{(i)} = \begin{bmatrix} x^{(i)}_1 \\ x^{(i)}_2 \end{bmatrix}$

$x^{(i)} = \begin{bmatrix} 3^{(i)} \\ 2^{(i)} \end{bmatrix}$

$x^{(i)} = 3$

$x^{(i)} = 0$

$x^{(i)} = 0$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left| y^{(i)} - \hat{y}^{(i)} \right|$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss

$\text{Error} = \frac{1}{n} \sum_{i=0}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]^2$

Error, Squared loss