Agenda :-

- Intro.
- Two Variables
- Multivariate Function
- Quadratic Function
- Cost function.
- Optimization
- [Evaluation Metric] - "R2 Score"

Intro

[Excel]

Formula

→ used case study

Formula "used for writing form"

enough "show by hand" ....

Home Work

Data

Fit ... to y

$x^{(1)} \to y^{(1)}$

$x^{(2)} \to y^{(2)}$

$x \to y$

"data"

$f(x) \to y$

Goal:

$y = f(x)$

$x \to y$

$x \in \mathbb{R}^d$

Supervised

x_{new} → [Every Fun] (unknown function) → $f(x)$ model → Predictions "Supervised" Price

y_{new} → Predicted Price.

$\frac{\partial f(x)}{\partial x} \approx \frac{\Delta f(x)}{\Delta x}$ "slope that"

$\hat{y}_{new} = \hat{y}_{\theta}$

[Training Fun] → $\{x^{(i)}, y^{(i)}\}_{i=1}^{n}$

Two Phases

① Training Phase.

vs/g (hidden function) → we find/fit the function $f(x)$

② Testing phase :

Use → error ϵ_n → Improve → Model → $\hat{y} \approx y$

Geometric Interpretation.

① "age" → Price

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

$\hat{y} = w_1 x + w_0$ ← $\mathbb{R}^d$ of value.

$y \approx w_1 x + w_0$

$y \approx mx + c$

1 feature → 2 dim

② "age", "edumaks" → Price

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

$\hat{y} = w_1 x_1 + w_2 x_2 + w_0$

x_1, x_2 → Phase.

2 features → 3dim

"3 dim for model"

d-features ⇒ $(d+1)$ dim hyp. to "model" the model.

$\hat{y} = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + w_0$

$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}_{(d,1)}$ $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}_{(d,1)}$

$\hat{y} = \underbrace{[w_1 \; w_2 \; \dots \; w_d]}_{1 \times d} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}_{(d,1)} + w_0$

$\hat{y} = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + w_0$

$\boxed{y = w^T x + w_0}$ → d-dim

Intuition

$\hat{y} = w_1 x + w_0$

$w_1 = +1$ $w_0 = 5$

power 3 -2

Find $w_1, w_2, w_3, \dots, w_d$ & w_0

"w" vector & w_0

Task : find Best w's

Univariate Linear Regression.

Use

$f(x) = w_1 x + w_0$

$y^{(1)} = f(x^{(1)})$

$y^{(1)} - \hat{y}^{(1)} = \epsilon^{(1)}$

"residuals"

$\left\{ \underset{w_1, w_0}{\arg\min} \sum_{i=1}^{n} \epsilon^{(i)^2} \right\}$ → find w_1, w_0 while minimizing the residuals.

Power sum of errors

$\epsilon^{(1)} = y^{(1)} - \hat{y}^{(1)}$
$= +5$

$\epsilon^{(2)} = y^{(2)} - \hat{y}^{(2)}$
$= -3$

Error/loss ⇒ $\sum_{i=1}^{n} \epsilon^{(i)}$

⇒ $0 + 5 + (-3) + (-1) + 0 + (-3)$

⇒ 0

⇒ Positive & negative errors cancel out each other.

MAE

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

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

"Mean Squared Error"

MSE

$$\boxed{\text{loss function } (L) = \frac{1}{n} \sum_{i=1}^{n} \left(y^{(i)} - (w_1 x^{(i)} + w_0) \right)^2}$$

MSE

Optimization.

$\underset{w_1, w_0}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left[y^{(i)} - (w_1 x^{(i)} + w_0) \right]^2$

→ Can be solved using GD/SGD

$[w_1, w_0]$

$w_1 = 0$

$w_0 = 5$

w_1

w_0

w_1

w_0

$[w_1, w_0]$

$[(w_1 - 0)]^2$

$w_1 = 0$

$[(w_1 - (w_1 - 5)]^2 = 2$

$w_1 = 5$

$[(w_1 - 5)]^2$

$w_1 = 5$

$\frac{\partial L}{\partial w_1} = 0 \to 4w_1^2 L$

$\frac{\partial L}{\partial w_0} = 0 \to 4w_0^2 L$

$\frac{\partial L}{\partial w_1} = 0 \to 4w_1^2 L$

$\frac{\partial L}{\partial w_0} = 0 \to 4w_0^2 L$

$\frac{\partial L}{\partial w_1} = 0 \to 4w_1^2 L$

$\frac{\partial L}{\partial w_0} = 0 \to 4w_0^2 L$

$\left.\begin{array}{l} \text{Solve} \\ \text{for } w_1, w_0 \end{array}\right\}$?

Impossible

Practical

Gradient Descent

w_1^{old}, w_0^{old}

① Randomly initialize (w_1, w_0)

② Repeat until convergence

$w_1^{new} = w_1^{old} - \eta \left[\frac{\partial L}{\partial w_1^{old}} \right]$

$w_0^{new} = w_0^{old} - \eta \left[\frac{\partial L}{\partial w_0^{old}} \right]$

→ Low learning rate α

Max. iteration 20,000

$w_1^{old} = 2.5000$
$w_0^{old} = 2.5000$

Gradients

$$L = \frac{1}{n} \sum_{i=1}^{n} \left[y^{(i)} - (w_1 x^{(i)} + w_0) \right]^2$$

$$L^{(i)} = \left[y - \underbrace{(w_1 x + w_0)}_{z} \right]^2$$

$\frac{\partial L^{(i)}}{\partial w_1} = \frac{\partial z^2}{\partial z} \cdot \frac{\partial z}{\partial w_1}$

$\Rightarrow 2z \cdot \frac{\partial [y - w_1 x - w_0]}{\partial w_1}$

$\Rightarrow 2[y - \hat{y}] \cdot (-x)$

$\Rightarrow 2[y - \hat{y}]$

$$\boxed{\frac{\partial L}{\partial w_1} \Rightarrow \frac{2}{n} \sum_{i=1}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right]}$$

$\frac{\partial L^{(i)}}{\partial w_0} = \frac{\partial z^2}{\partial z} \cdot \frac{\partial z}{\partial w_0}$

$\Rightarrow 2z \cdot \frac{\partial [y - w_1 x - w_0]}{\partial w_0}$

$\Rightarrow 2[y - \hat{y}] \cdot (-1)$

$\Rightarrow 2[y - \hat{y}]$

$$\boxed{\frac{\partial L}{\partial w_0} \Rightarrow \frac{2}{n} \sum_{i=1}^{n} \left[y^{(i)} - \hat{y}^{(i)} \right] \cdot x}$$