# CSE 427 : Machine Learning

Quiz → 15%
Assign → 15%
Mid → 20%
Final → 50%

(E) → (T) → (P)

Experience Task Performance

| Training Data | Testing Data |
| --- | --- |
| 70% | 30% |
| 80% | 20% |

i) Supervised Learning

ii) Unsupervised Learning.

iii) Reinforcement Learning.

Nobel Prize Name ⇒
Mathematics = Fields Medal
Comp. Sci. = ACM AM Turing Award.

Dr. Holly Kreiger
Youtube Channel ⇒ Number file

# আমার কাছে 5L ও 3L এর Jar আছে। আমাকে 4L পানি মেপে দিতে হবে।

⇒ ১/ 5L ভরব
২/ 5L থেকে 3L এ ঢালব (5L এ 2L থাকবে)
৩/ 3L খালি করব।
৪/ 5L এর মধ্যে যে 2L আছে সেটা 3L এ ঢালব।
৫/ 5L ভরব।
৬/ 5L থেকে 3L এর বোতল পুরোটা ভরব।
(3L এর মধ্যে আগে 2L ছিল, এখন আরও 1L ভরলো)
৭/ ∴ 5L এ এখন 4L আছে।

## Supervised Learning

- Regression (value পাবো)
- Classification (Class পাবো)

1000 mail এর মধ্যে 850 → non-spam, 150 → spam } Classification

Weather Data
- → আগামীকাল Temp. কত হবে? → Regression
- → আগামীকাল বৃষ্টি হবে কি হবেনা? → Classification.

1000 image data
- → Male/Female → Classification
- → Age → Regression.

## Single Variable Linear Regression

১টা feature থাকবে।

① Dataset

| (Feature) $x$ | $y$ (Output) |
|---|---|
| 0 | 4 |
| 1 | 7 |
| 2 | 7 |
| 3 | 8 |

(II) Hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 * x \qquad y = mx + c$$

(III) Parameter Initialization:

feature $n$ সংখ্যক হলে parameter $n+1$ সংখ্যক হবে।

$\theta_0, \theta_1 \rightarrow$ parameter

ধরি, $\theta_0 = 2, \theta_1 = 2$

(Randomly initialize করা Dataset range এর ভিতর করব)

$$h_\theta^{(1)}(x) = \theta_0 + \theta_1 * x^{(1)}$$
$$= 2 + 2 * 0 = 2$$

$$h_\theta^{(2)}(x) = \theta_0 + \theta_1 * x^{(2)}$$
$$= 2 + 2 * 1 = 4$$

$$h_\theta^{(3)}(x) = \theta_0 + \theta_1 * x^{(3)}$$
$$= 2 + 2 * 2 = 6$$

$$h_\theta^{(4)}(x) = \theta_0 + \theta_1 * x^{(4)}$$
$$= 2 + 2 * 3 = 8$$

(IV) cost function: (ভুলের পরিমাণ/error বের করব)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left( h_\theta^{(i)}(x) - y^{(i)} \right)^2$$

$$= \frac{1}{2 \times 4} \left\{ (2-4)^2 + (4-7)^2 + (6-7)^2 + (8-8)^2 \right\}$$

$$= \frac{1}{8} (4+9+1+0)$$

$$= \frac{1}{8} \times 14$$

$$= 1.75 \text{ error/cost.}$$

Next Step এ $\theta_0, \theta_1$ এমনভাবে change করব যাতে এই error কম আসে

(V) Gradient Descent:

repeat until convergence

$$\left\{ \underbrace{\theta_j}_{\text{নতুন}} = \overbrace{\theta_j}^{\text{পুরাতন}} - \boxed{\frac{\overbrace{\alpha}^{\text{Learning rate.}}}{\underbrace{m}_{\text{\# of data entry}}} \frac{d}{d\theta_j} \{ J(\theta) \}} \right\}$$

$$\theta_0 := \theta_0 - \frac{\alpha}{m} \boxed{\sum_{i=1}^{m} \left( h_\theta^{(i)} - y^{(i)} \right)}$$

$$\theta_1 := \theta_1 - \frac{\alpha}{m} \boxed{\sum_{i=1}^{m} \left( h_\theta^{(i)} - y^{(i)} \right) \cdot x^{(i)}}$$

Let, $\alpha = 0.01$ (ছোট সংখ্যা ধরব)

$$\theta_0 := 2 - \frac{0.01}{4}\{(2-4)+(4-7)+(6-7)+(8-8)\}$$

$$= 2 - \frac{0.01}{4}\{-2-3-1+0\}$$

$$= 2.015$$

$$\theta_1 := 2 - \frac{0.01}{4}\{(2-4)\cdot 0+(4-7)\cdot 1+(6-7)\cdot 2+(8-8)\cdot 3\}$$

$$= 2 - \frac{0.01}{4}\{0-3-2+0\}$$

$$= 2.0125$$

এবার নতুন $\theta_0, \theta_1$ দিয়ে Hypothesis function বের করব। cost function বের করব। আবার $\theta_0, \theta_1$ update করব।

এভাবে, $h_\theta(x) = \theta_0 + \theta_1 * x$

↳ $x$ এর যেকোনো value-র জন্য

→ $y$ এর prediction পাব।

Plot: $x$ এর আপেক্ষে $h$

$x$ এর আপেক্ষে $y$.

# Multi variable Linear Regression

(i) Dataset:

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 1 | 0.5 | 2 |
| 1 | 1.5 | 3 |
| 2 | 1 | 4 |
| 3 | 1 | 4 |

(ii) $\theta_0, \theta_1, \theta_2$    Let, $\theta_0 = 1, \theta_1 = 0.5, \theta_2 = 1$

(iii) Hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2$$

$$h_\theta^{(1)}(x) = \theta_0 + \theta_1 * x_1^{(1)} + \theta_2 * x_2^{(1)}$$

$$= 1 + 0.5 * 1 + 1 * 0.5 = 2$$

$$h_\theta^{(2)}(x) = \theta_0 + \theta_1 * x_1^{(2)} + \theta_2 * x_2^{(2)}$$

$$= 1 + 0.5 * 1 + 1 * (1.5)$$

$$= 3$$

$$h_\theta^{(3)}(x) = \theta_0 + \theta_1 * x_1^{(3)} + \theta_2 * x_2^{(3)}$$

$$= 1 + 0.5 * 2 + 1 * 1$$

$$= 3$$

$$h_\theta^{(4)}(x) = \theta_0 + \theta_1 * x_1^{(4)} + \theta_2 * x_2^{(4)}$$

$$= 1 + 0.5 * 3 + 1 * 1$$

$$= 3.5$$

## (iv) Cost Function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta^{(i)} - y^{(i)}\right)^2$$

$$= \frac{1}{2\times 4} \left\{ (2-2)^2 + (3-3)^2 + (3-4)^2 + (3.5-4)^2 \right\}$$

$$= \frac{1}{8}(0+0+1+0.25) = \frac{1.25}{8} = 0.156225$$

## (v) Gradient Descent:

Repeat until convergence:

$$\left\{ \theta_{j\in\{0,1,2\}} := \theta_j - \frac{\alpha}{m} \frac{\partial}{\partial \theta_i} \{J(\theta)\} \right.$$

$$\}$$

$$\theta_0 := \theta_0 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)}(x) - y^{(i)}\right)$$

$$\theta_1 := \theta_1 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)}(x) - y^{(i)}\right) \cdot x_1^{(i)}$$

$$\theta_2 := \theta_2 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)}(x) - y^{(i)}\right) \cdot x_2^{(i)}$$

$$\therefore \theta_0 := 1 - \frac{0.01}{4} \left\{ (2-2) + (3-3) + (3-4) + (3.5-4) \right\}$$

$$= 1 - \frac{0.01}{4} \left\{ 0 + 0 + (-1) + (-0.5) \right\}$$

$$= 1.00375$$

$$\theta_1 := 0.5 - \frac{0.01}{4}\{(2-2)1+(3-3)\cdot 1+(3-4)\cdot 2+(3.5-4)\cdot 3\}$$

$$= 0.5 - \frac{0.01}{4}\{0+0+(-2)+(-1.5)\}$$

$$= 0.50125$$

$$\theta_2 := 1 - \frac{0.01}{4}\{(2-2)(0.5)+(3-3)(1.5)+(3-4)\cdot 1+(3.5-4)\cdot 1\}$$

$$= 1 - \frac{0.01}{4}\{0+0+(-1)+(0.5)\}$$

$$= 1.00375$$

# Dataset Vectorization:

$\boxed{x_1, x_2, x_3 \ldots x_{100}}$ → feature set/vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{100} \end{bmatrix} \rightarrow \text{feature vector} \qquad x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \\ \vdots \\ x_{100}^{(1)} \end{bmatrix} \qquad x^{(3)} = \begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \\ x_3^{(3)} \\ \vdots \\ x_{100}^{(3)} \end{bmatrix}$$

যে matrix এর row n-সংখ্যক, column 1টা, তারে Matrix কে বলবে vector. $\underset{\text{row}}{n} \times \underset{\text{colm.}}{1}$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{100} \end{bmatrix} \qquad h_\theta(x) = \theta^T \cdot x$$

$$= \begin{bmatrix} \theta_0, \theta_1, \theta_2 \cdots \theta_{100} \end{bmatrix} \begin{bmatrix} x_0 = 1 \\ x_1 \\ x_2 \\ \vdots \\ x_{100} \end{bmatrix} 101 \times 1$$

$$= \theta_0 \cdot x_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \cdots + \theta_{100} \cdot x_{100}$$

$$= \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \cdots + \theta_{100} \cdot x_{100}$$

* for solving large number of Dataset it can be done in dot গুণ which is vectorization.

## Normalization:

| $x_1$ | $x_2$ | $x_3$ |
|---|---|---|
| 0.2 | 105 | 15280 |
| 0.7 | 270 | 17570 |
| 0.125 | 380 | 19390 |
| 0.42 | 170 | 2010 |

Normalization কখন করব?

⇒ যখন Database এর feature এ বিভিন্ন range এর value থাকবে।

## Mean Normalization:

$$x_1^{(1)} := \frac{x_i - \mu_i}{s_i}$$

$$= \frac{0.2 - 0.36125}{0.575}$$

$$= -0.2804$$

$$\mu_i(x_1) = \frac{0.2 + 0.7 + 0.125 + 0.42}{4}$$

$$= 0.36125$$

$$s_i(x_1) = \max(x_i) - \min(x_i)$$

$$= 0.7 - 0.125$$

$$= 0.575$$

$$x_2^{(1)} := \frac{105 - 231.25}{275}$$

$$= -0.4590$$

$$\mu_i(x_2) = \frac{105 + 270 + 380 + 170}{4}$$

$$= 231.25$$

$$s_i(x_2) = 380 - 105$$

$$= 275$$

[-1, 1] এর মধ্যে থাকবে।

$$x_3' = \frac{15280-13562.5}{17380}$$
$$= 0.0988$$

$$\mu_i(x_3) = \frac{15280+17570+19390+2010}{4}$$
$$= 13562.5$$

$$S_i(x_3) = 19390 - 2010$$
$$= 17380$$

এভাবে প্রত্যেক $x_1, x_2, x_3$ এর জন্য update value বের করা লাগবে।

Normalization এর জন্য $\alpha$ not needed; $\alpha$ এর value normalization এর উপর dependent.