Recall in Lec 1 the example of calculating final score (0~100) under supervised learning.

The type of supervised learning used was called regression.

Ex.

Given the training data below, what would the score be ($y$) if the student studied 7 hours ($x$)?

| $x$ (hours) | $y$ (score) |
|---|---|
| 10 | 90 |
| 9 | 80 |
| 3 | 50 |
| 2 | 30 |

training data → Regression Model

$x = 7$ → Regression Model → $y = ?$

Let us look at a simplified example:

| X | Y |
|---|---|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |

This is clearly $y = x$ pattern (linear)

∴ Our initial hypothesis for predicting the y-value given an $x$-value would be to use/find a line of best fit (This is called linear regression)

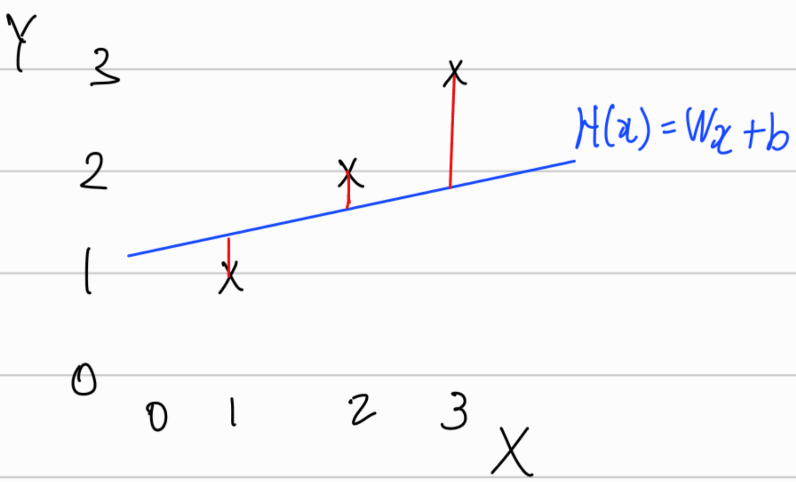
$\therefore$ Set Hypothesis $H(x) = Wx + b$

and find $W$ and $b$ values

But how do we find good values for $W$ and $b$ when we have a less simple data set?

A: use a cost/loss function.

Cost function example: $H(x) - y$




Note that $H(x)-y$ can be positive or negative at different points. Which means large positive disparities might cancel out large negative disparities when summed.

$\therefore$ New Cost function idea, $(H(x)-y)^2$

2 Benefits to squaring:

1. Removes negatives (no cancellation of opposite sign cost)
2. Large disparities become larger and smaller

ones become smaller. i.e. $10^2 = 100$, $0.1^2 = 0.01$

A more formal calculation would look like:

$$\frac{(H(x_1)-y_1)^2+(H(x_2)-y_2)^2+(H(x_3)-y_3)^2}{3}$$

This gets the average of the disparities squared.

$\therefore cost = \frac{1}{m}\sum_{i=1}^{m}(H(x_i)-y_i)^2$ where $m$ is the number of data points

Since $H(x) = Wx+b$, we can express cost as a function of $W$ and $b$.

$$cost(W,b) = \frac{1}{m}\sum_{i=1}^{m}(H(x_i)-y_i)^2$$

$\therefore$ The ultimate goal is to minimize this cost, and the process of finding $W$ and $b$ values that minimize the cost is "learning"

학습목표: $\underset{W,b}{\text{minimize}}\ cost(W,b)$