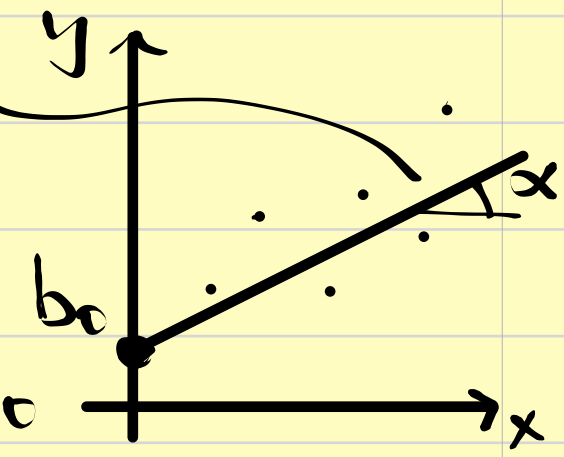
Regression Algorythms: Linear, Polynomic, Logistic

Hypothesis: we have data from the system.

Linear Reg. $y = b_0 + b_1 x$



$b_1 \equiv tg\alpha$

$b_0 \equiv$ cross with $x=0$

Step 1. Gather the data

	x	y
cw 1	3	6'5
cw 2	4	5'5
cw 3	6	13
cw 4	3	3'5

Step 2. Mean value of the variables.

**The linear regression crosses the mean value

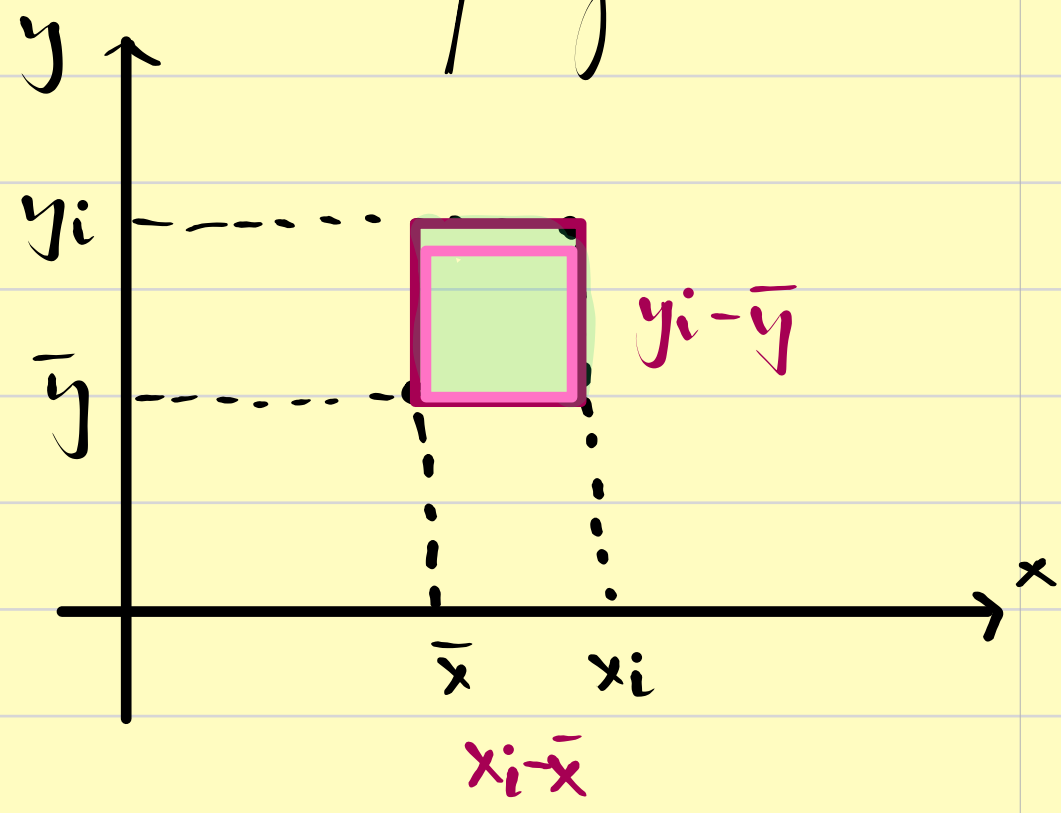
$$\bar{y} = b_0 + b_1 \bar{x}$$

$$\bar{x} = \frac{1}{N}\sum x_i = \frac{1}{4}\left[3+4+6+3\right] = 4$$

$$\bar{y} = \frac{1}{N}\sum y_i = \frac{1}{4}\left[6'5+5'5+13+3'5\right] = 7'87$$

Step 3. Calculate the slope of the line

$$b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$



$$b_1 = \frac{(3-4)(6'5-7'57)+(4-4)(8'5-7'57)+(6-4)(13-7'57)+(3-4)(3'5-7'57)}{(3-4)^2+(4-4)^2+(6-4)^2+(3-4)^2} = 4'53$$

$$y = b_0 + 4'53\,x$$

Step 4. b_0.

$$\bar{y} = b_0 + 4'53\,\bar{x} \rightarrow 7'57 = b_0 + 4'53 \cdot 4 \rightarrow b_0 = -11'46$$

$$\hat{y} = -11'46 + 4'53\,x$$

If $x = 3'9$, what is the predicted value of y?

$$\hat{y} = -11'46 + 4'53 \cdot 3'9 = 7'39 .$$

Polynomic Reg.

Quadratic Reg. $y = ax^2 + bx + c$

x	y
1	2'5
2	5'8
3	11'9
4	21'4
5	31'2

$$2'5 = a \cdot 1^2 + b \cdot 1 + c \quad (1)$$
$$5'8 = a \cdot 2^2 + b \cdot 2 + c \quad (2)$$
$$11'9 = a \cdot 3^2 + b \cdot 3 + c \quad (3)$$

$(2)-(1)$: $5'8 - 2'5 = 4a - a + 2b - b + c - c$

$$3'3 = 3a + b \quad (4)$$

$(3)-(2)$: $11'9 - 5'8 = 9a - 4a + 3b - 2b + c - c$

$$6'1 = 5a + b \quad (5)$$

$(5)-(4)$: $6'1 - 3'3 = 5a - 3a + b - b \rightarrow a = 1'4$

$\xrightarrow{(4)} b = -0'9 \xrightarrow{(1)} c = 2$

$$\hat{y} = 1'4x^2 - 0'9x + 2$$

What is the value of $\hat{y}$ if $x = 0$?

$$\hat{y}(x=0) = 2$$

Logistic Regression.

Is a type of Regr. used when the dependent variable (y) is binary.

It predicts the probability of the outcome based on one or more variables (x_i)

We use a threshold (i.e 0'5) to decide if the output is 0 or 1.

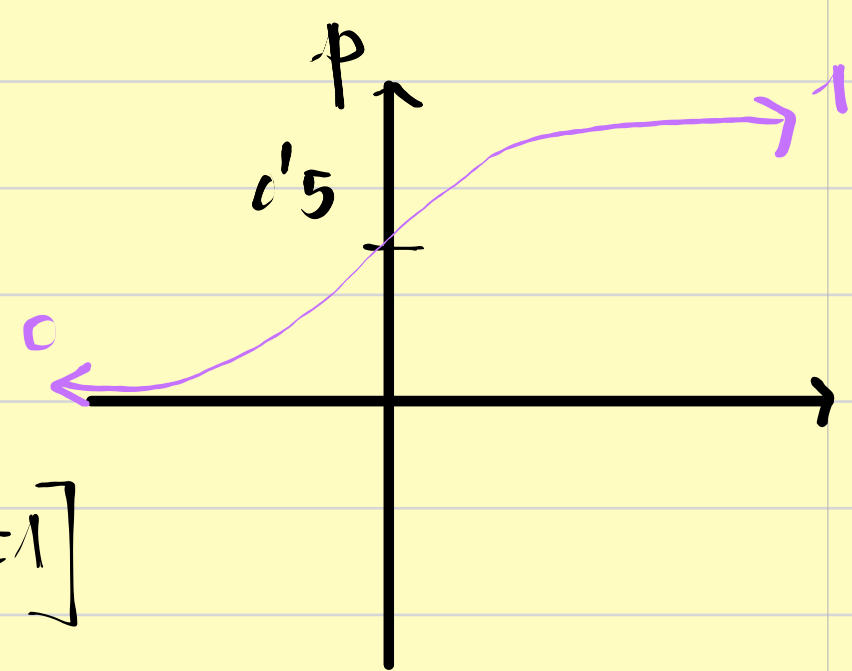
Example: A professor wants to predict wether a student will pass ($\hat{y}=1$) or fail ($\hat{y}=0$) an exam based on the number of hours they study (x).

Step 1. Gather data

x[h]	y[Pass[1], No pass[0]]
1	0
2	0
3	0
4	1
5	1
6	1

Step 2. Use the Logit Function to model the relationship.

$$\ln\left[\frac{p}{1-p}\right] = b_0 + b_1 x$$



p: probability of pass [y=1]
b_0: is the intercept x=0
b_1: slope of x

$$\frac{p}{1-p} = e^{b_0+b_1 x} \rightarrow p = e^{b_0+b_1 x} - p\left[e^{b_0+b_1 x}\right]$$

$$\rightarrow p\left[1+e^{b_0+b_1 x}\right] = e^{b_0+b_1 x} \rightarrow p = \frac{1}{1+e^{-(b_0+b_1 x)}}$$

Step 3. Solve.

$x=3 \rightarrow y=0$

$x=4 \rightarrow y=1$

**we assume $x=4 \rightarrow p=0'5 \rightarrow \ln\left[\frac{0'5}{1-0'5}\right] = b_0+b_1 \cdot 4$

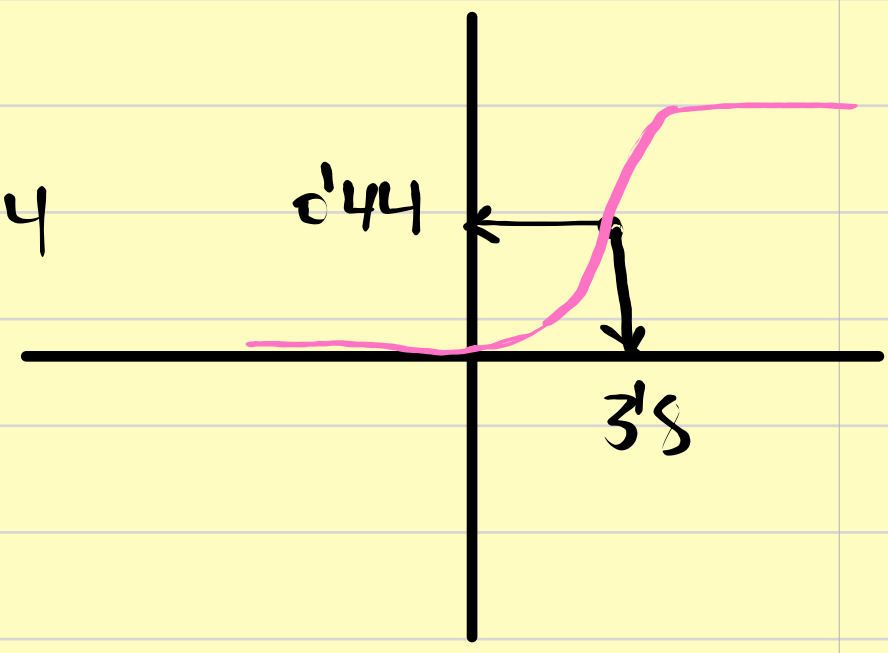
" " $x=6 \rightarrow p=0'9 \rightarrow \ln\left[\frac{0'9}{1-0'9}\right] = b_0+b_1 \cdot 6$

$$\ln\left[\frac{1}{1-1}\right] = 0 = b_0 + b_1 \cdot 4$$

$$\ln\left[\frac{0.9}{0.1}\right] = b_0 + b_1 \cdot 6$$

$$\ln\left[\frac{0.9}{0.1}\right] = 2b_1 \rightarrow b_1 = 1.1$$

$$0 = b_0 + 1.1 \cdot 4 \rightarrow b_0 = -4.4$$



$$p = \frac{1}{1+e^{-(-4.4+1.1x)}}$$

What is the prob of pass after 3.8 h study?

44%.