Logistic Regression Algorithm.

* Logistic regression is a data analysis techniques that uses mathematics to find the relationships b/w 2 data factors.
* Predict one factor based on other

Eg. certain People will buy a Computer or not.

→ find possible classifiers
↳ fix decision surface.
↓
← data point
(buy a computer / not)

→ beyond classification
* Interested in Accuracy (PAC qn)
→ Not just in the right classification

→ Eg. Medical domain
If patient have specific disease walk to hospital to meet doctor. check / tells the doctor confidence level, if they says 95% sure that this patient is affected by specific diseases doctor

will start treatment.

- Confidence of classification.

→ Treat regression Problem

* Use an indicator variable for class.

customer buy Comp - 1

" Not " - 0

$x_1 = \langle 30000, 25 \rangle$ · y_1 = Does not buy Comp

$x_2 = \langle 80000, 45 \rangle$ · y_2 = buys Computer

⋮

* Use linear regression

→ Linear regression is not limited in range.

- Output cannot be interpreted as Probability

- Can be -ve.

→ Transformed function will use this linear regression.

- Logistic or logit function
  - Log - odds
  - Let $p(x)$ denote $p(y=1/x)$
  - Logit transformation is given by

$$\log\left(\frac{p(x)}{1-p(x)}\right)$$

→ fitting linear regression

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + x\cdot\beta_1$$

Solving $p(x)$ - sigmoid

$$p(x) = \frac{e^{\beta_0 + x\cdot\beta_1}}{1+e^{\beta_0 + x\cdot\beta_1}} = \frac{1}{1+e^{-(\beta_0 + x\cdot\beta_1)}}$$

* Not suitable for all kind of problems
* Predict class is 1 when $p(x) > 0.5$ &
  0 otherwise
  - Minimizes misclassification rate.
* linear classifier
  Decision boundary @ $\beta_0 + x\cdot\beta_1 = 0$
* powerful