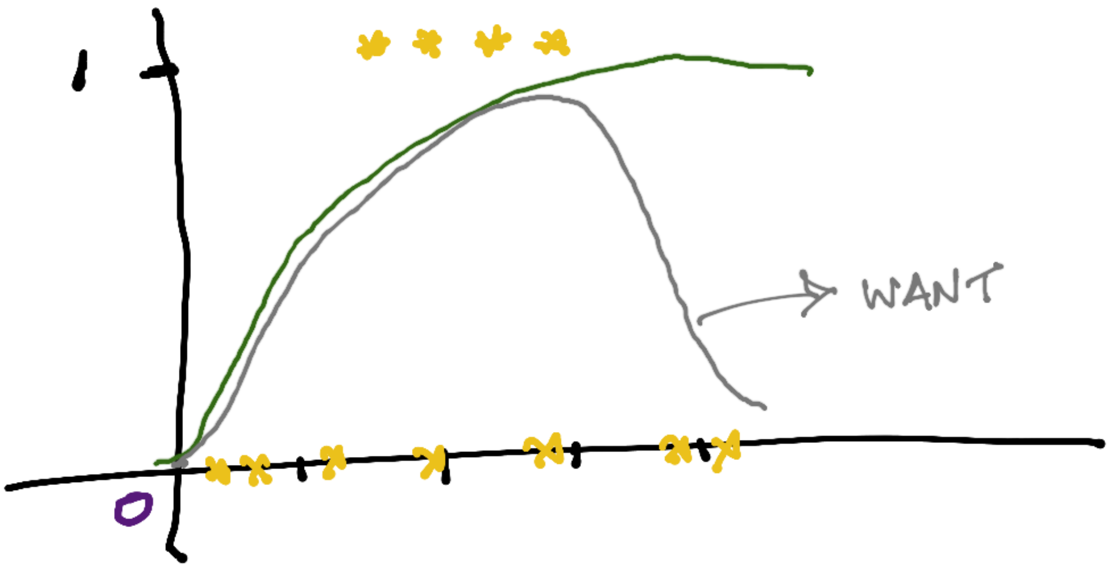
Regression & Classification

24th Oct 2019



$$y = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$$

$$\log_e \frac{p}{1-p} = \beta_0 + \beta_1 x \Rightarrow p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Monotic function. Hence p increases as x increases.

Try.
$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x + \beta_2 x^2$$

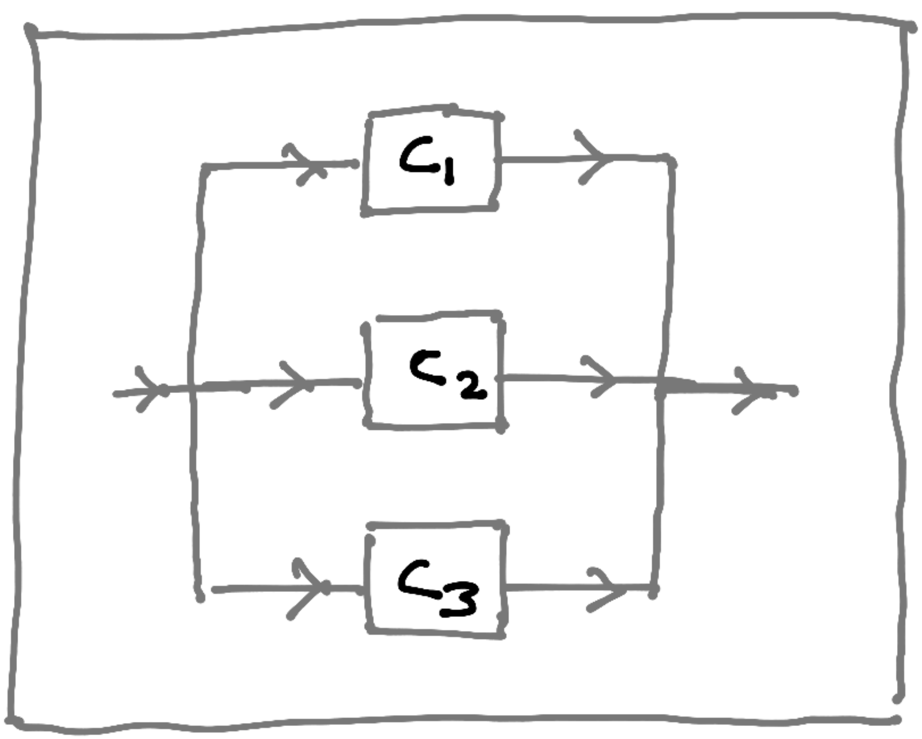
$$\log \frac{p}{1-p} = \beta_0 + \beta_1 \sin(x) + \beta_2 \cos(x)$$

Deep Learning does automated feature engineering.

For a car manufacturer, what is the probability that the system will break down?

Failure is extremely rare. Hence traditional classification models won't work.

Here the data is <u>very</u> skewed.



D	C_1	C_2	C_3	S
1	0	1	0	0
2	0	0	1	0
⋮	⋮	⋮	⋮	⋮
n	0	0	0	0

Separate components may fail. But the entire system never fails.

Can create probability distns for $\underbrace{c_1, c_2, c_3}_{\text{independent}}$

$$P[c_1, c_2, c_3] = P[c_1]P[c_2]P[c_3]$$

But the probability should also be a function of time.

Prob of failure in Day 1 < Prob of failure in Day n

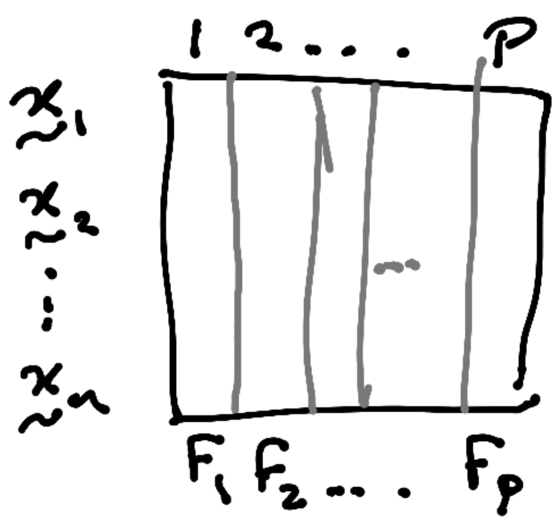
$P[c_1(t)]$ Not just a probability distn. It's a process [stochastic]

$$P[c_1(t), c_2(t), c_3(t)] = P[c_1(t)]\,P[c_2(t)]\,P[c_3(t)]$$

Most popular → Non homogeneous Poisson Process.

Simulate many scenarios & check how many times failures occure.

Copula Distribution



$$\Sigma_{n \times p}$$

$$F(x_1, x_2, \ldots x_p) = c(F_1, F_2, \ldots, F_p, \Sigma)$$

Creates Proper Joint Distn