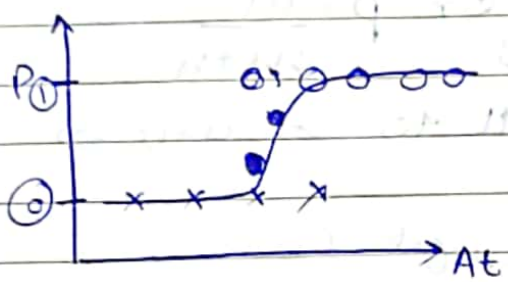
CS229: Logistic regression.

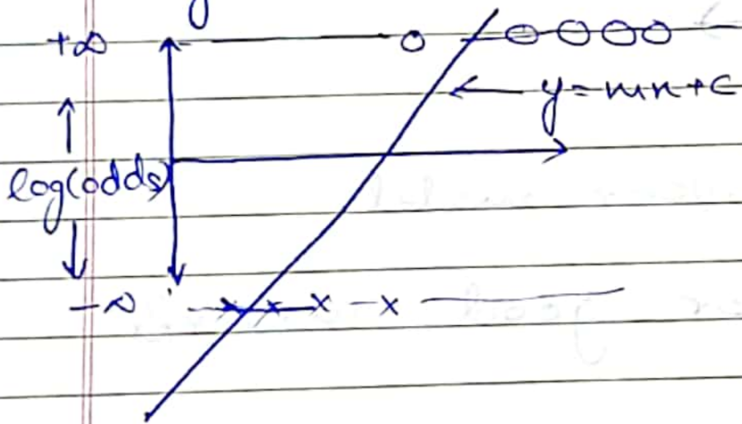
- Odds = Probability in favour / Probab against
- Likelyhood = Product of probability of truth.

Probability curve →

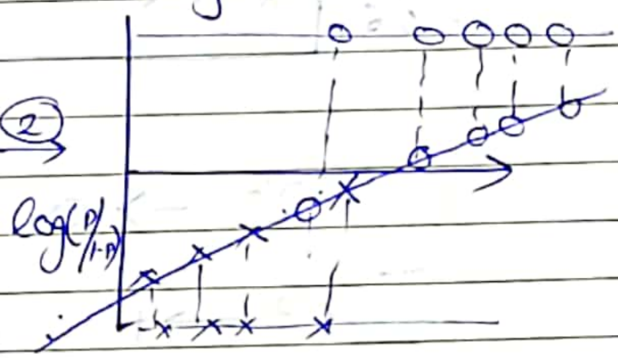


①

Log odds curve

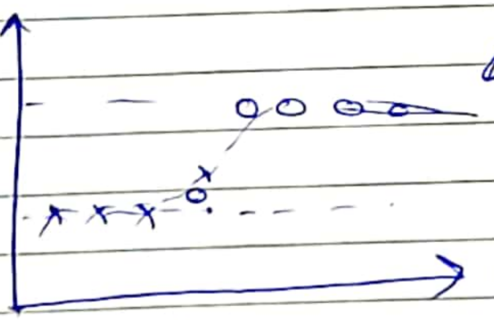


Project



$$P = \frac{e^{\log(odds)}}{1+e^{\log(odds)}}$$

Redraw Probability ③



likelyhood =

$$\prod_{i=\text{all } o} P(\text{is } o \mid o) \times \prod_{i=\text{all } x} P(\text{is } x \mid x)$$

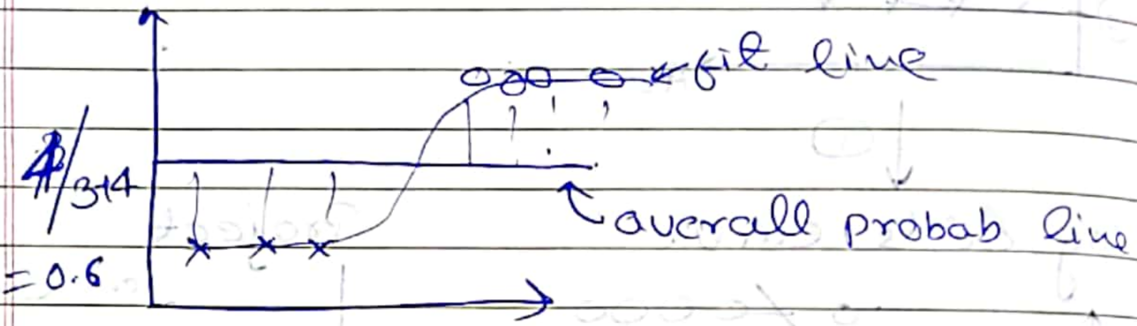
④

log(likelyhood) → 0 for good fit.

$$R^2 = \frac{\text{log likely(overall p)} - ll(\text{fit})}{ll(\text{overall } p)}$$

ll(overall p) → this is the line drawn for $p = \frac{m}{m+n}$

(m is the =1) ∥ to x axis ⇒



$R^2 \to 0$ poor model

$R^2 \to 1$ for good models

2(LL(fit) − LL(overall probab))

$$p(y|x;\theta) = h_\theta(x)^y \times (1-h_\theta(x))^{1-y}.$$

Likelyhood of a point can hence be:

$$L(\theta) = \prod_{i=1}^{n} \left[h_\theta(x^{(i)})\right]^{y^{(i)}} \left[1-h_\theta(x^{(i)})\right]^{1-y^{(i)}}$$

$$\ell(\theta) = \log[L(\theta)] = \cancel{\#}$$

$$= \sum_{i=1}^{n} \left(y^{(i)} \log(h_\theta(x^{(i)})) + (1-y^{(i)}) \cdot \log(1-h_\theta(x^{(i)})) \right)$$

Gradient Ascent

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = \left[y \times \frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)} \right] \times \frac{d}{d\theta} g(\theta^T x)$$

Simplifies to $\rightarrow$

$$= (y - h_\theta(x))\, x_j$$

$$\theta_j := \theta_j + \alpha\,(y^{(i)} - h_\theta(x^{(i)}))\, x_j^{(i)}.$$