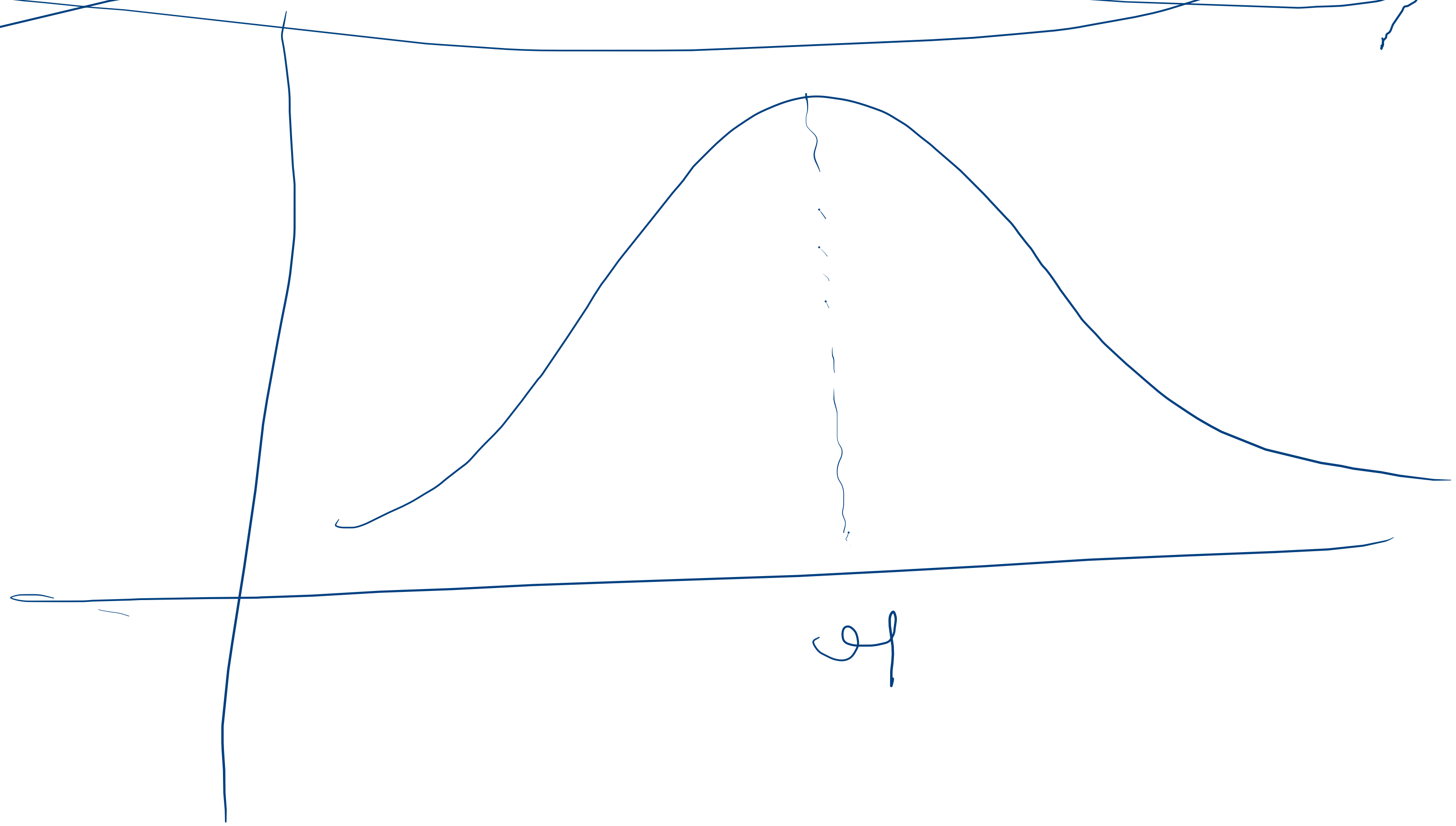
$$p(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$\rightarrow \quad (x_1, y_1), (x_2, y_2), \ldots \ldots (x_n, y_n)$

$\rightarrow \quad (\mu, \sigma)$

$$(y_1, y_2, \ldots y_n)$$

$$\ln(p(?)) = -\ln(\sigma\sqrt{2\pi}) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

↓

cost for a single data point

the summation over loss func. would be the summation of costs for all the data pnts.

$$L_{(\mu,\sigma)} = -N\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2}$$

to find μ & σ

do $\frac{\partial L}{\partial \mu} = 0$ & $\frac{\partial L}{\partial \sigma} = 0$

$$\ln(p(?)) = -\ln(\sigma\sqrt{2\pi}) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

$\downarrow$

cost for a single data point

our loss func. would be the summation of costs for all the data pnts.

$$L_{(\mu, \sigma)} = -N\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2}$$

to find μ & σ

do $\frac{\partial L}{\partial \mu} = 0$ & $\frac{\partial L}{\partial \sigma} = 0$

$$\frac{\partial L}{\partial \mu} = \frac{\cancel{2}\sum_{i=1}^{n}(x_i - \mu)}{\cancel{2}\sigma^2} = 0$$

$$\sum x_i - N\mu = 0 \Rightarrow \boxed{\mu = \frac{\sum x_i}{N}}$$

$$\frac{\partial L}{\partial \sigma} = \frac{-N}{\sigma\cancel{\sqrt{2\pi}}} \times \cancel{\sqrt{2\pi}} + \frac{\cancel{2}\sum_{i=1}^{n}(x_i - \mu)^2}{\cancel{2}\sigma^3} = 0$$

$$\frac{N}{\sigma} = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{\sigma^3} \Rightarrow \boxed{\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

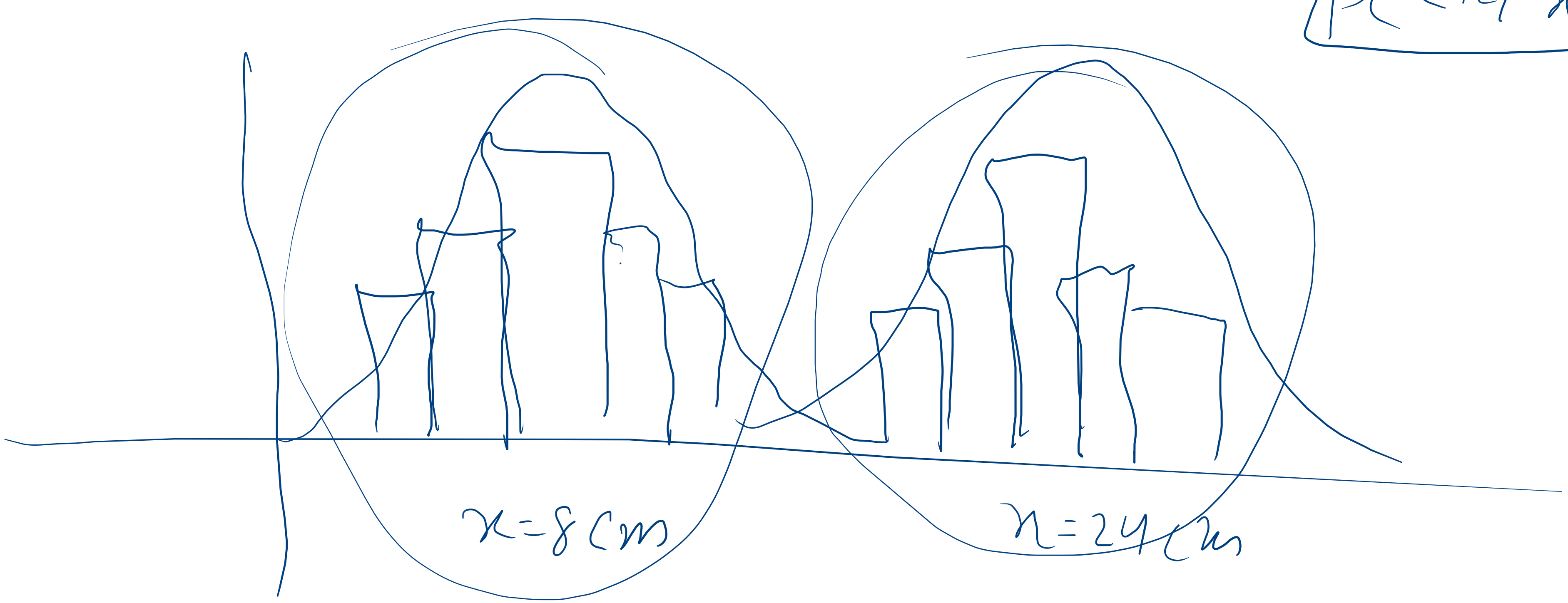
Male & Female.

$x \rightarrow$ hair length.

→ all features are indep.

→ each feature follows it's own gauss.

$\boxed{P(C_k / x)}$



$$P(C_k \mid x) = \frac{P(x \mid C_k)\, P(C_k)}{P(Data)}$$

posterior

likelihood

prior prob. of the class.

→ Input vec. x

→ K pre-defined classes

→ $P(C_k / x)$ max. among them for the right class.

$$P(x_1, x_2, \ldots x_K \mid \theta)$$

Logistic :—

→ Binary classif.

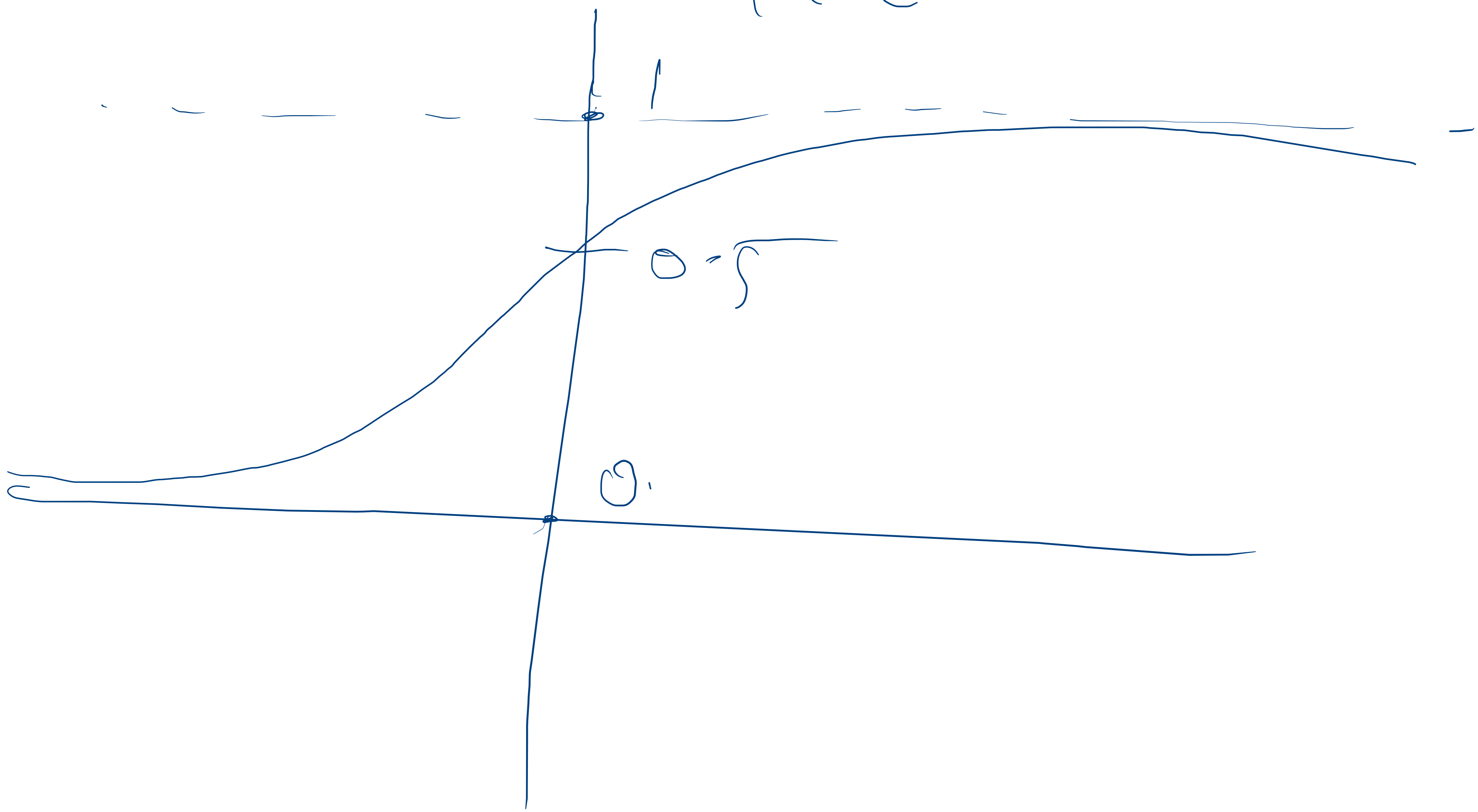
1 → +ve class

0 → −ve class.

Build a relationship to find this prob.
for input(x).

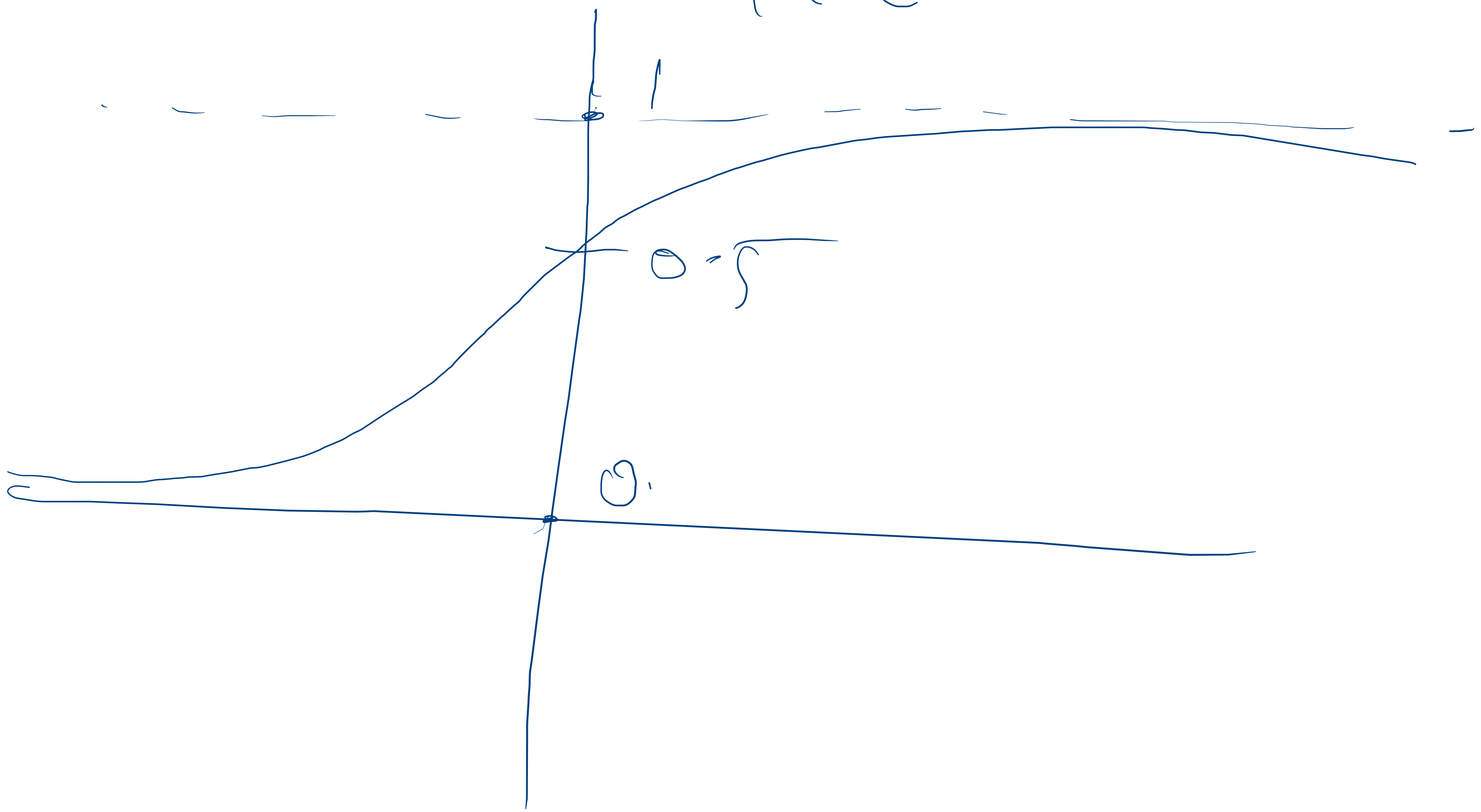
Sigmoid function :-

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



Sigmoid function :-

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



i) first train a line

$$y = w^T x$$

ii) Add non-linearity

$$\sigma(y) = \frac{1}{1 + e^{-\underline{w}^T x}}$$

parameters

belonging to class 1.

i) first train a line

$$y = w^T x$$

ii) Add non-linearity

$$\sigma(y) = \frac{1}{1 + e^{-w^T x}}$$

parameters

belonging to class 1.

i) first train a line

$$y = w^T x$$

ii) Add non-linearity

$$\sigma(y) = \frac{1}{1 + e^{-w^T x}}$$

parameters

belonging to class 1.

Likelihood :-

We know we have only 2 possible outputs
0 & 1

$1 \longrightarrow h(x)$

$0 \longrightarrow 1 - h(x)$

$$= \prod_{i \mid y_i = 1} h(x) \times \prod_{i \mid y_i = 0} (1 - h(x))$$

taking -ve log likelihood.

$$L = \left(\left(-\sum_{i \mid y_i = 1} \ln(h(x_i)) \right) + \left(-\sum_{i \mid y_i = 0} \ln(1 - h(x_i)) \right) \right)$$

$$= \sum_{i=1}^{N} y_i \ln(h(x_i)) + (1 - y_i) \ln(1 - h(x_i))$$

$$h(x_i) = \frac{1}{1 + e^{-w^T x_i}}$$

optimal parameter.

MLE

$$h_w(x_i) = \frac{1}{1+e^{-w^T x_i}}$$

h_w

$$\boxed{h_w(x_i) = \frac{1}{1+e^{-w^T x_i}}}$$

$$L_w = -\sum_{i=1}^{N} \left[y_i \log\left(\frac{1}{1+e^{-w^T x_i}}\right) + (1-y_i)\log\left(1-\frac{1}{1+e^{-w^T x_i}}\right)\right]$$

$$\boxed{L_w = \sum_{i=1}^{N} \left[y_i \log\left(1+e^{-w^T x_i}\right) + (1-y_i)\log\left(1+e^{w^T x_i}\right)\right]}$$

$$\frac{\partial L}{\partial \omega} = \sum_i \left(y_i \times \frac{1}{1+e^{-\omega^T x_i}} \times -x_i \times e^{-\omega^T x_i} \right.$$

$$\left. + \left((1-y_i) \times \frac{1}{1+e^{\omega^T x_i}} \times x_i \times e^{\omega^T x_i} \right) \right)$$

$$= \sum_i \frac{-y_i x_i e^{-\omega^T x_i} + (1-y_i)\, x_i}{1+e^{-\omega^T x_i}}$$

$$= \sum_i \frac{x_i - x_i y_i \left(1+e^{-\omega^T x_i}\right)}{1+e^{-\omega^T x_i}}$$

$$= \sum_i x_i \left(\frac{1}{1+e^{-w^T x_i}} \right) - x_i y_i$$

$$= \sum_i x_i \left(\sigma(w^T x_i) - y_i \right)$$

$$\frac{\partial L}{\partial w} = \boxed{\underbrace{X^T \left[\sigma(w^T X) - Y \right]}_{\text{Matrix form.}}}$$

$$\boxed{W_{new} = W_{old} - \alpha \left(\frac{\partial L}{\partial w} \right)}$$