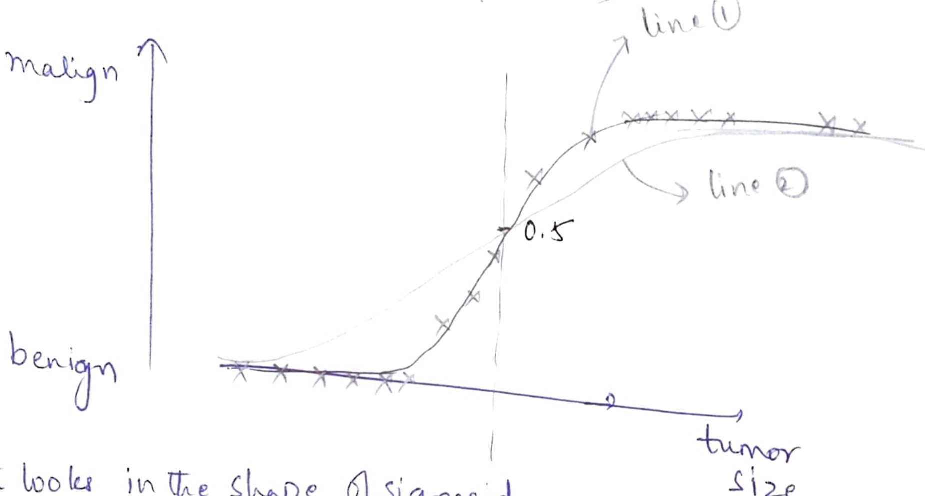
30/03/2024 **Lecture. 21**

* In linear regression, extrapolation is difficult & tricky.

Logistic Regression:

↳ actually, it's a 2-class classification algorithm

Say, given data looks something like this;



→ Data looks in the shape of sigmoid.

→ Sigmoid function is called logistic function;

$$\frac{1}{1+e^{-(w_1x+w_0)}}$$

→ 2 dimensions

→ Put a logit function & put threshold = 0.5

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

$> 0.5 \Rightarrow$ malign

$< 0.5 \Rightarrow$ benign.

↓ applies for any dimension but here we take 2 dimensions only.

We can write cost function like this;

$$\sum \left(\sigma(w^Tx_i - y_i)^2\right)$$

↓ but, we cannot do this because this is not convex.

$p(y_i = 1 | x_i, w) = [\sigma(w^T x_i)]^{y_i}$

Using MLE,

$p(y_i = 0 | x_i, w) = [1 - \sigma(w^T x_i)]^{1-y_i}$

$\prod_{i=1}^{N} (\sigma(w^T x_i))^{y_i} \cdot (1 - \sigma(w^T x_i))^{1-y_i}$

This function will be higher for line ①, lower for line ②

Take log on b.s,

$J(w) = \sum_{\forall i} y_i \log(\sigma(w^T x_i)) + (1-y_i) \log(1 - \sigma(w^T x_i))$

$J'(w) = \sum \frac{y_i}{\cancel{\sigma(w^T x_i)}} \left(\cancel{\sigma(w^T x_i)}(1 - \sigma(w^T x_i))(x_i)\right) - \frac{1-y_i}{\cancel{(1-\sigma(w^T x_i))}} \left(\sigma(w^T x_i)\cancel{(1-\sigma(w^T x_i))} x_i\right)$

$= \sum y_i (1 - \sigma(w^T x_i))(x_i) - (1-y_i)(\sigma(w^T x_i))(x_i)$

$= \cancel{\sum y_i (1 - \sigma(w^T x_i))}$

$\left[(y_i - y_i \cancel{\sigma(w^T x_i)}) - \sigma(w^T x_i) + \cancel{y_i \sigma(w^T x_i)}\right] x_i$

$J'(w) = \sum (y_i - \sigma(w^T x_i)) x_i$

You can apply gradient descent here