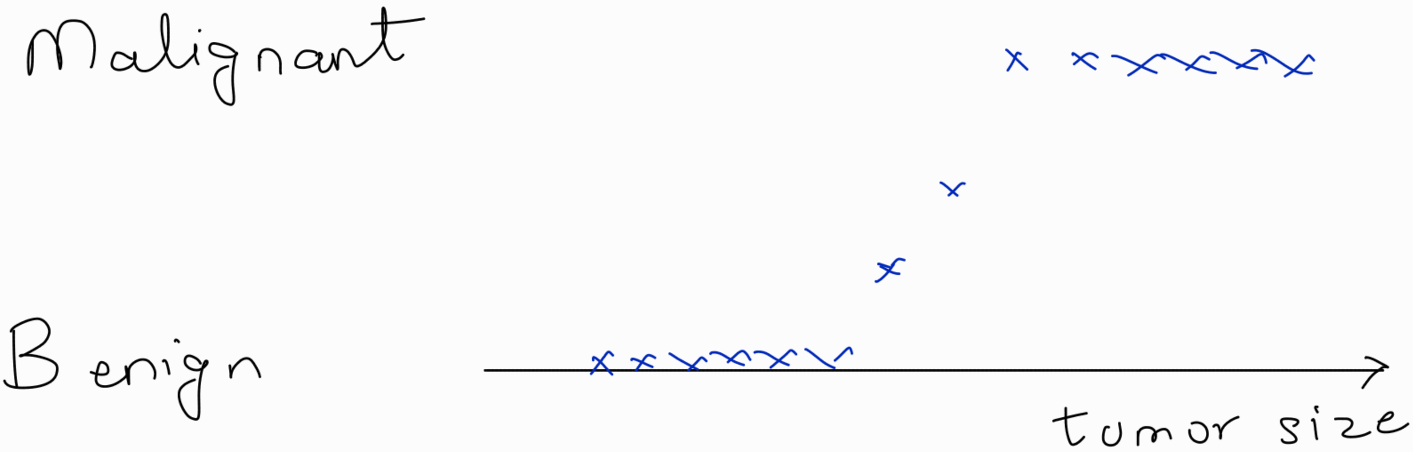
30/3/24

The sigmoid function is logistic regression



$$\sigma(w^T x) = \frac{1}{1 + e^{-(w_1 x + w_0)}}$$

It fits logit function but is actually a binary classifier.

Gives the probability of a data point belonging to each class.

Set a threshold above which you

predict particular class

Loss Function :

We guess the MSE

$$L = \sum \left[\sigma(w^T x_i) - y_i\right]^2$$

But this is non-convex. We find MLE

$$P(y_i = 1 \mid x_i, w) = \left[\sigma(w^T x_i)\right]^{y_i}$$

$$P(y_i = 0 \mid x_i, w) = \left[(1 - \sigma(w^T x_i)\right]^{(1-y_i)}$$

$$\prod_{i=1}^{N} \left[\sigma(w^T x_i)^{y_i} \cdot \left[(1 - \sigma(w^T x_i)\right]^{(1-y_i)}\right.$$

Taking log,

$J(W) =$

$$\sum_{\forall i} y_i \log\left[\sigma(W^T x_i)\right] + (1-y_i)\log\left[1-\sigma(W^T x_i)\right]$$

$J'(W)$

$$= \sum \frac{y_i}{\sigma(W^T x_i)}\left[\sigma(W^T x_i)(1-\sigma(W^T x_i)\right] x_i$$

$$- \frac{(1-y_i)}{\cancel{\left[1-\sigma(W^T x_i)\right]}}\left[\sigma(W^T x_i)\cdot\cancel{(1-\sigma(W^T x_i))}\, x_i\right]$$

$$= \sum\left[y_i - y_i\sigma(W^T x_i) - \sigma(W^T x_i) + y_i\sigma(W^T x_i)\right] x_i$$

$$\therefore \boxed{J'(W) = \sum\left[y_i - \sigma(W^T x_i)\right] x_i}$$

We are maximizing $J'(W)$, gradient ascent, not descent.

$$\therefore W' = W + J'(W)$$

This is because $J(W)$ is not the error term, but the joint probability and we want to maximize this.