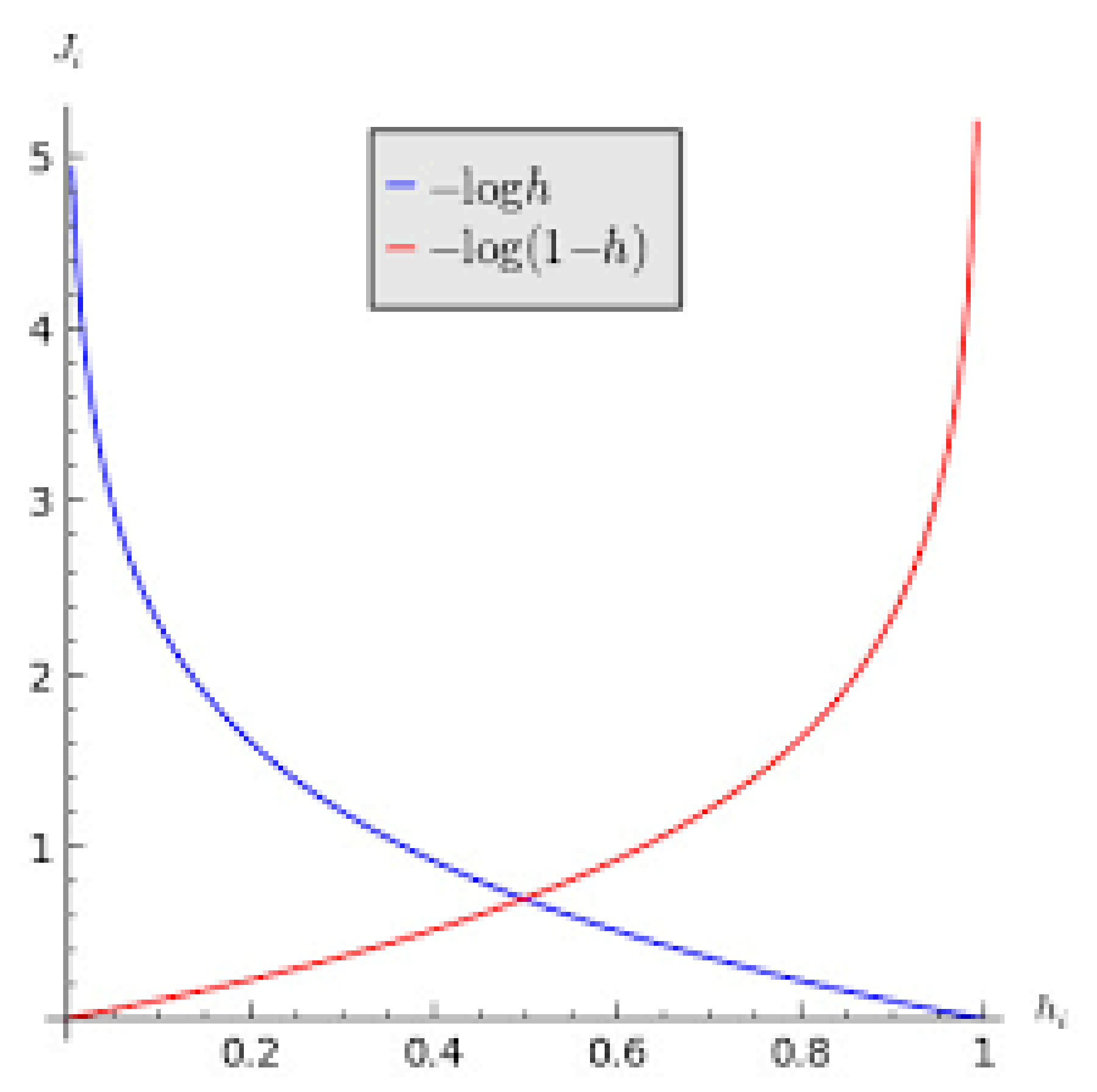
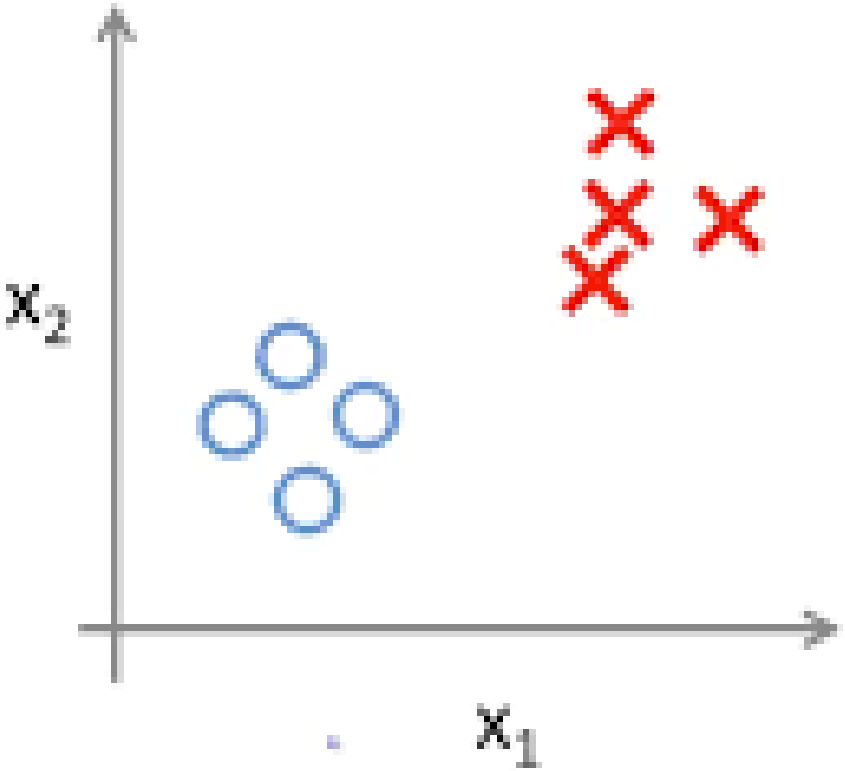
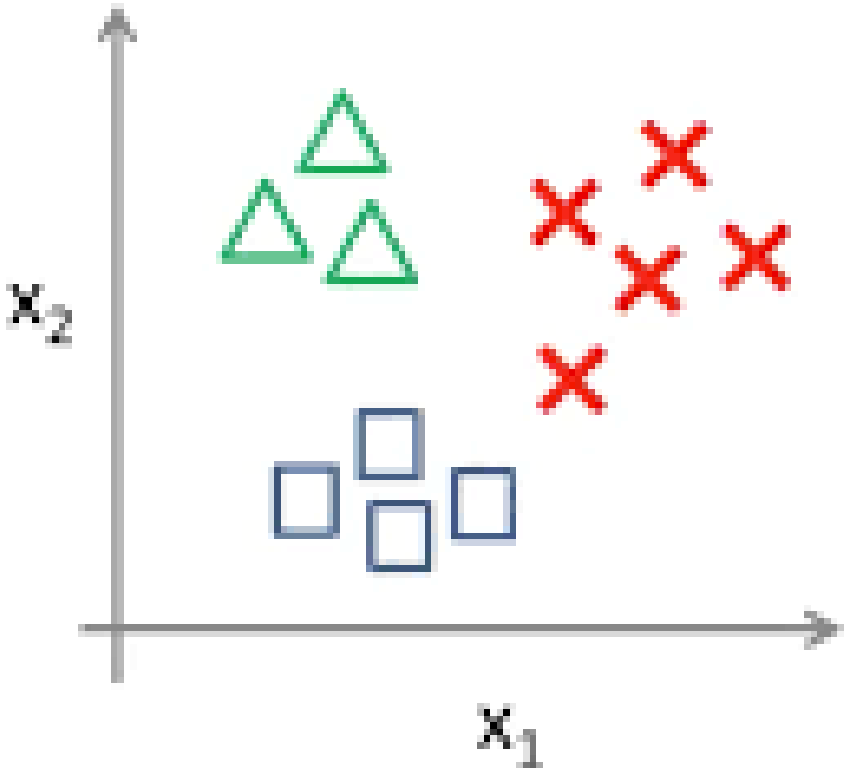

J_i
$-\log h$
$-\log(1-h)$
5
4
3
2
1
0.2
0.4
0.6
0.8
1
h_i

Binary classification:

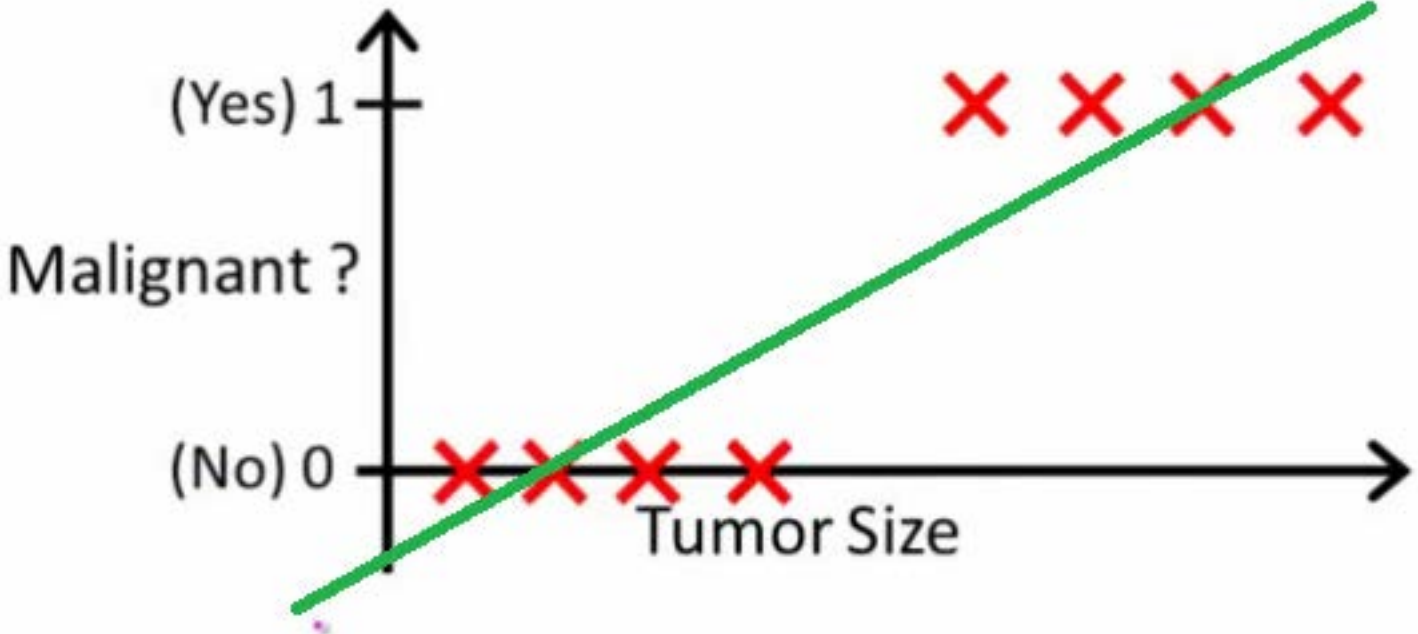


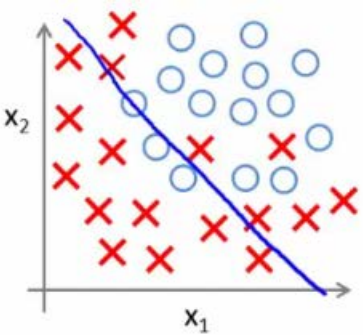
Multi-class classification:



$$h(x) = \frac{1}{1 + e^{-mx}}$$

only variable which needs to be optimized to get lowest cost

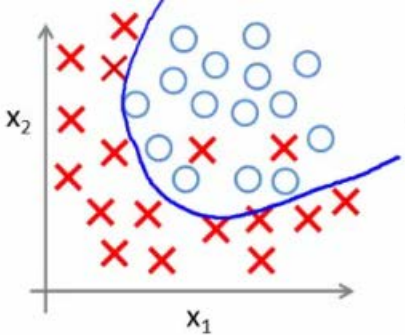

(Yes) 1
Malignant ?
(No) 0
Tumor Size



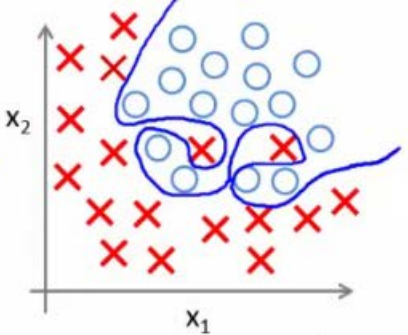
$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

(g = sigmoid function)

UNDERFITTING
(high bias)



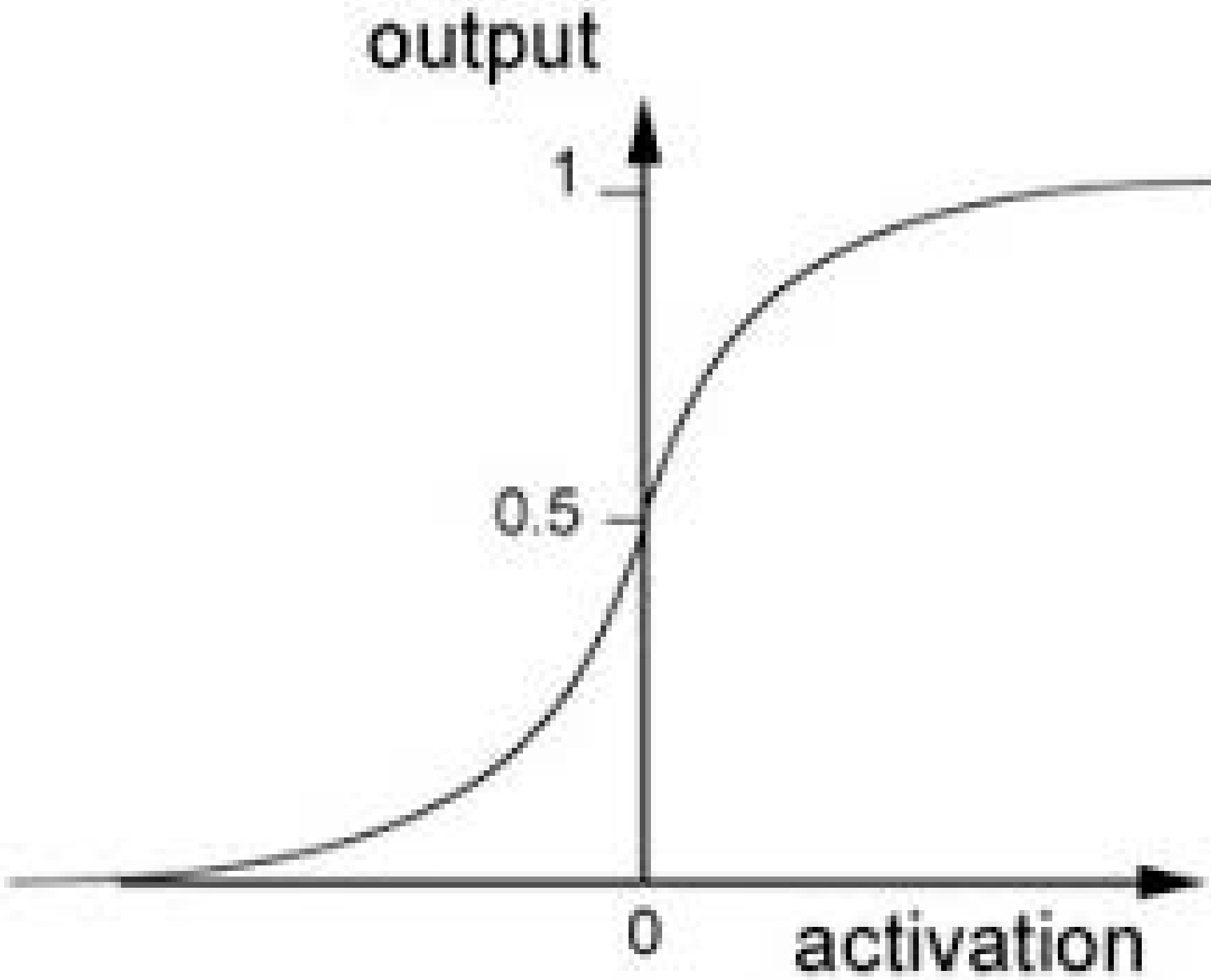
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$



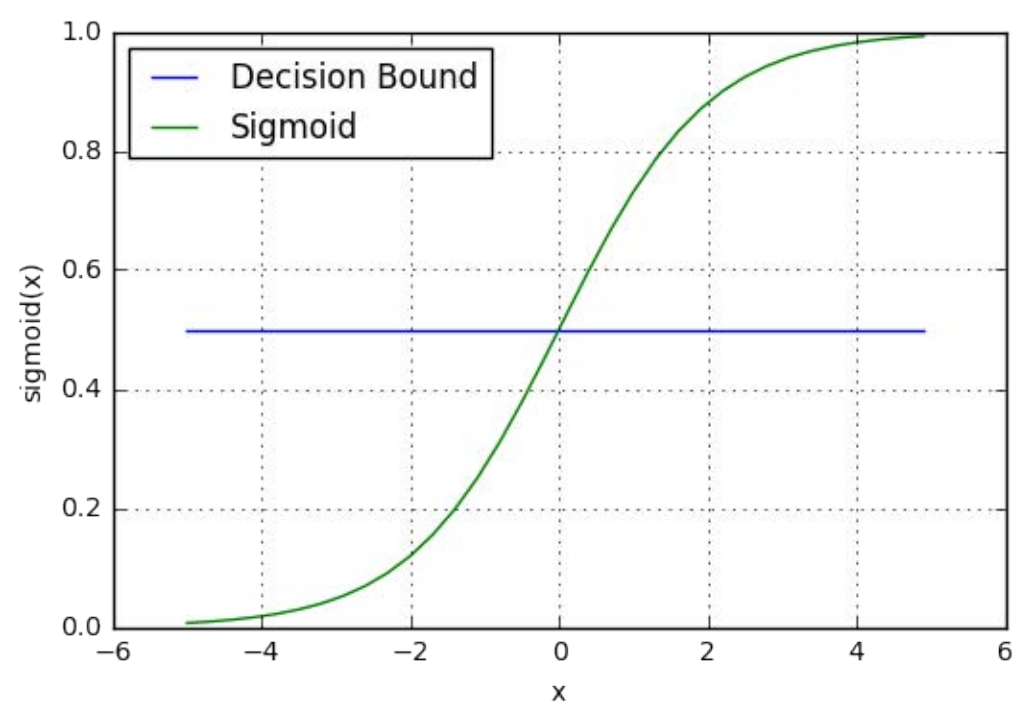
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots$

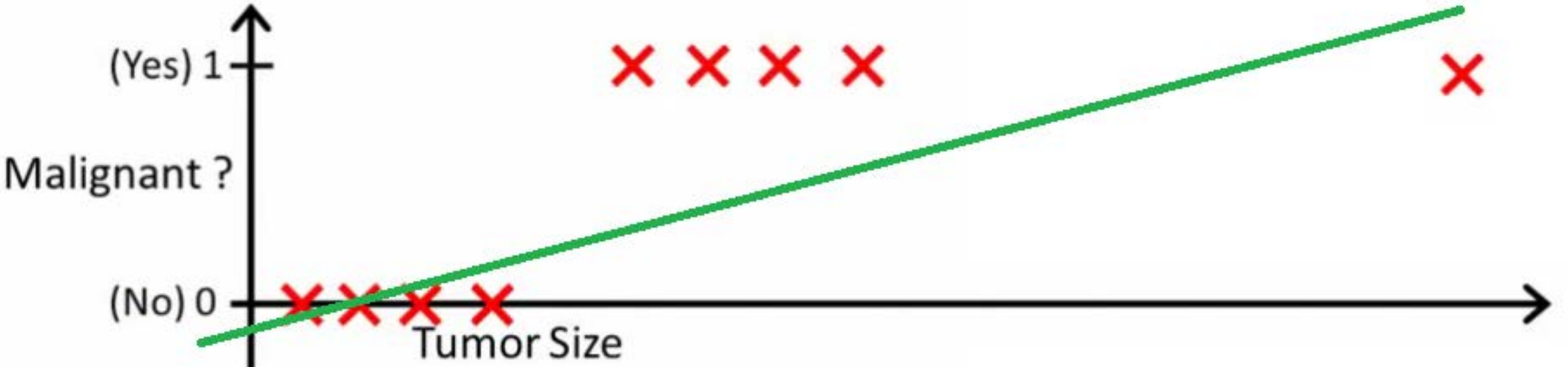
OVERFITTING
(high variance)

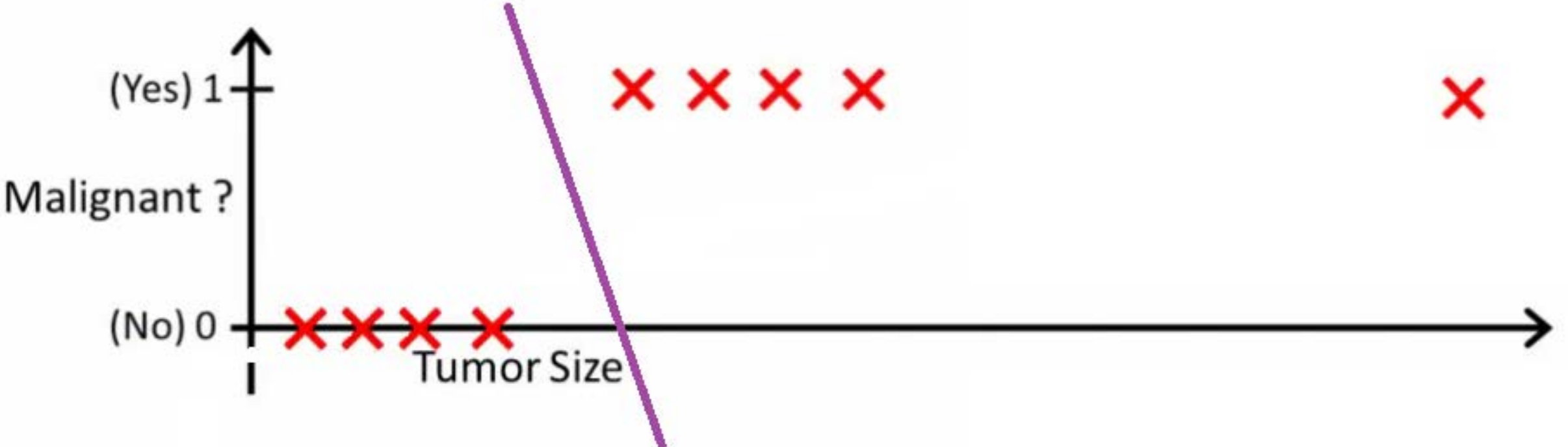
$$S(t) = \frac{1}{1 + e^{-t}}.$$


output
1
0.5
0
activation

$$S'(t) = S(t)(1 - S(t)).$$


Decision Bound
Sigmoid
sigmoid(x)
x
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6


(Yes) 1
Malignant ?
(No) 0
Tumor Size


(Yes) 1
Malignant ?
(No) 0
Tumor Size