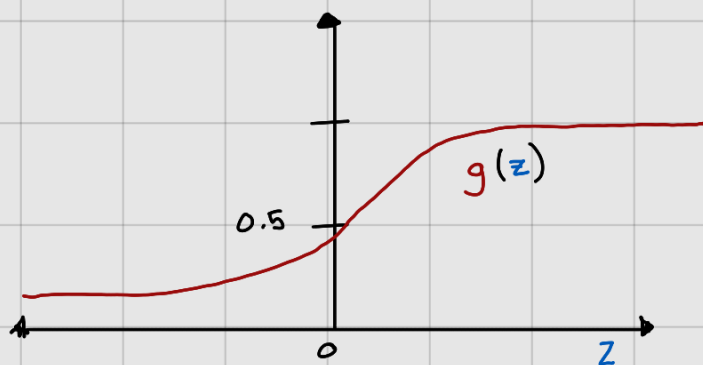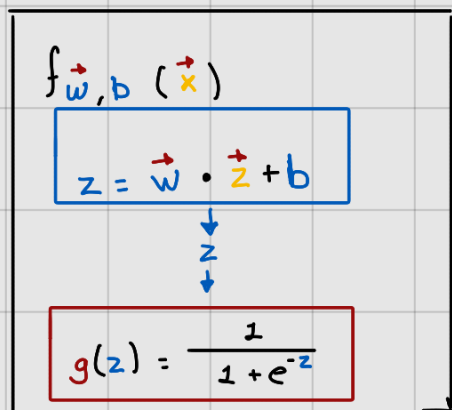# Logistic Regression

sigmoid function

logistic function

outputs between 0 and 1

$$g(z) = \frac{1}{1 + e^{-z}} \qquad 0 < g(z) < 1$$

$$f_{\vec{w},b}(\vec{x})$$

$$z = \vec{w} \cdot \vec{z} + b$$

$$z$$

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

$$f_{\vec{w},b}(\vec{x}) = g(\vec{w} \cdot \vec{z} + b) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

logistic Regression

Example

x is "tumor size"

y is 0 (not malignant)

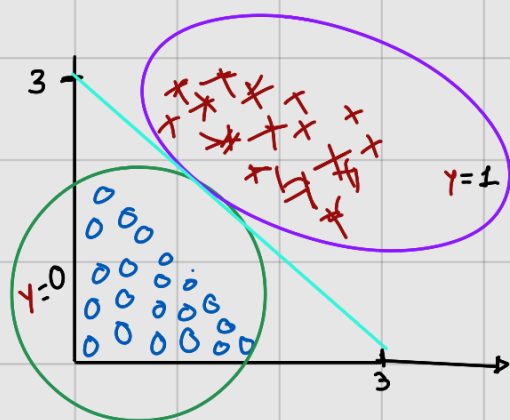or 1 (malignant)

$$f_{\vec{w},b}(\vec{x}) = 0.7$$

70% chance that y is 1

$$f_{\vec{w},b}(\vec{x}) = P(y = 1 | \vec{x} ; \vec{w}, b)$$

Probability that y is 1,

given input $\vec{x}$, parameters $\vec{w}, b$

Decision boundary

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underset{1}{w_1}x_1 + \underset{1}{w_2}x_2 + \underset{-3}{b})$$

Decision boundary

$$z = \vec{w} \cdot \vec{x} + b = 0$$
$$z = x_1 + x_2 - 3 = 0$$
$$x_1 + x_2 = \boxed{3}$$

Non-linear decision boundaries

Complex __ need Higher-order polynomial terms allow for more intricate decision boundaries

like: $f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2 + w_6 x_1^3 + \ldots + b)$

Cost function for logistic regression

cost
$$J(\vec{w},b) = \frac{1}{m} \sum_{i=1}^{m} L\underbrace{\left(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}\right)}_{loss}$$

New definition

$$L\left(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}(\vec{x}^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

convex
↳ can reach a

global minimum

Simplified loss function

$$L\left(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}\right) = \underbrace{-y^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)})) - (1 - y^{(i)}) \log(1 - f_{\vec{w},b}(\vec{x}^{(i)}))}$$

convex
(single global minimum)

$$J(\vec{w},b) = \frac{1}{m} \sum_{i=1}^{m} \left[L\left(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}\right) \right]$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)})) - (1 - y^{(i)}) \log(1 - f_{\vec{w},b}(\vec{x}^{(i)})) \right]$$

Gradient Descent Implementation in Logistic Regression

cost

$$J(\vec{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)})) - (1 - y^{(i)}) \log(1 - f_{\vec{w},b}(\vec{x}^{(i)})) \right]$$

repeat {
$\quad$ j = 1...n

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w}, b) \qquad \frac{\partial}{\partial w_j} J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{w}, b) \qquad \frac{\partial}{\partial b} J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})$$

} simultaneous updates

repeat {
$\quad$ j = 1...n

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) \right]$$

} simultaneous updates

Linear Regression $\qquad f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic Regression $\qquad f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$

$\underbrace{\qquad\qquad\qquad}_{\text{Sigmoid function}}$

Same concepts:

- Monitor gradient descent (learning curve)
- Vectorized implementation
- Feature scaling