



On Logistic Regression

Alexander G. Ororbia II
Introduction to Machine Learning
CSCI-335
10/7/2026

Special Thanks: Sargur N. Srihari, Andrew Ng, Carlos Guestrin

Machine Learning Algorithms

	Supervised Learning	Unsupervised Learning
Discrete	Classification K-NN classification	Clustering
Continuous	Regression Linear regression Polynomial regression K-NN regression	Dimensionality reduction Principal components analysis

Machine Learning Algorithms

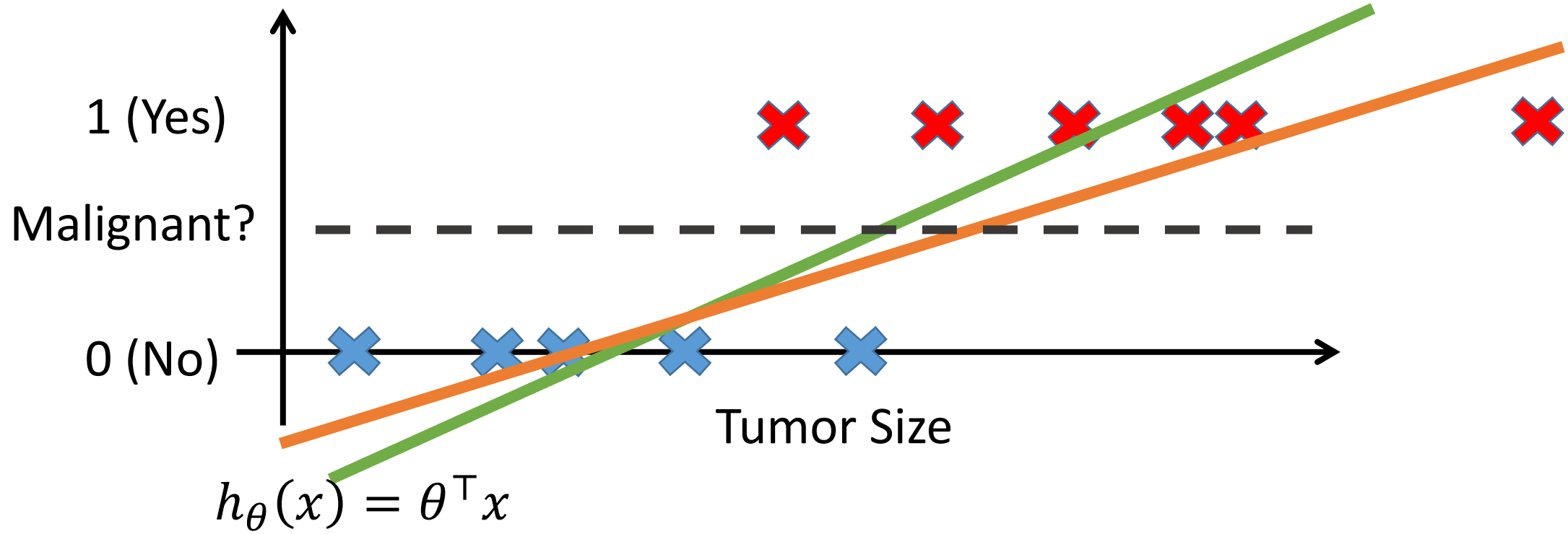
	Supervised Learning	Unsupervised Learning
Discrete	 Classification	Clustering
Continuous	  Regression	Dimensionality reduction

Logistic Regression

- Hypothesis *representation*
- Cost function (*evaluation*)
- Logistic regression with gradient descent (*optimization*)
- Regularization (again!)
- Multi-class classification

Logistic Regression

- **Hypothesis *representation***
- Cost function (*evaluation*)
- Logistic regression with gradient descent (*optimization*)
- Regularization (again!)
- Multi-class classification



- *Threshold* classifier output $h_{\theta}(x)$ at 0.5
 - If $h_{\theta}(x) \geq 0.5$, predict “ $y = 1$ ”
 - If $h_{\theta}(x) < 0.5$, predict “ $y = 0$ ”

Why use Logistic Regression?

- There are many important research topics for which the dependent variable is "limited" in some way
- *Example:* whether or not a person smokes/drinks/skips class/takes advanced mathematics
 - For these, outcome is neither continuous nor distributed normally
- **Binary** logistic *regression* is a type of regression analysis where dependent variable is a ***dummy*** variable:
coded 0 (*negative class*: did not smoke) or 1 (*positive class*: did smoke)

Categorization: $y = 1$ or $y = 0$

$h_{\theta}(x) = \theta^{\top} x$ (from linear regression)
can be > 1 or < 0

Logistic regression: $0 \leq h_{\theta}(x) \leq 1$

Logistic regression is actually for classification

Hypothesis Representation

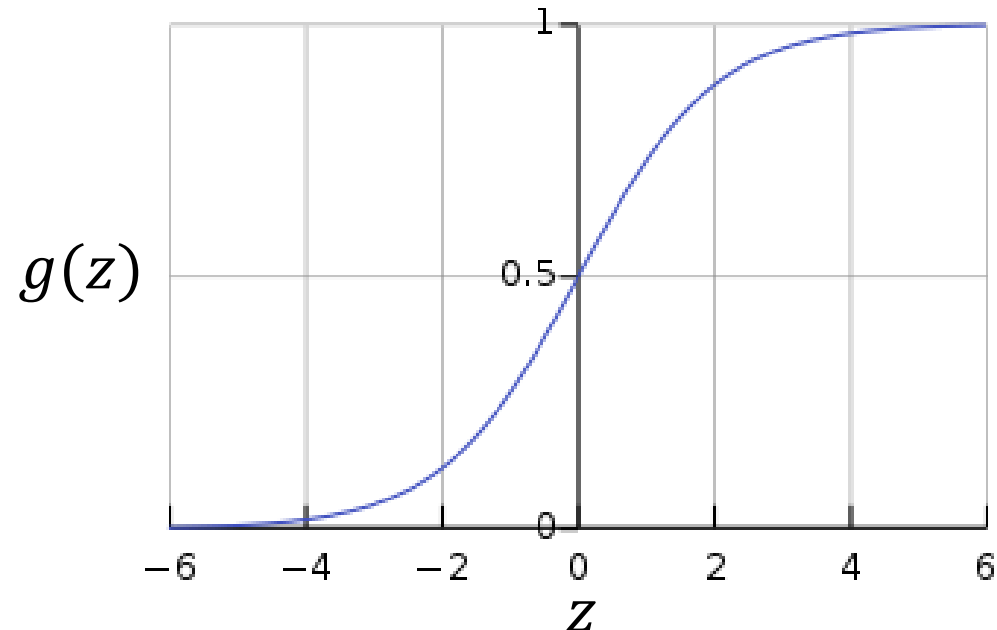
- Want $0 \leq h_{\theta}(x) \leq 1$

- $h_{\theta}(x) = g(\theta^{\top} x),$

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

- Sigmoid function
- Logistic (link) function

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^{\top} x}}$$



Interpretation of Hypothesis Output

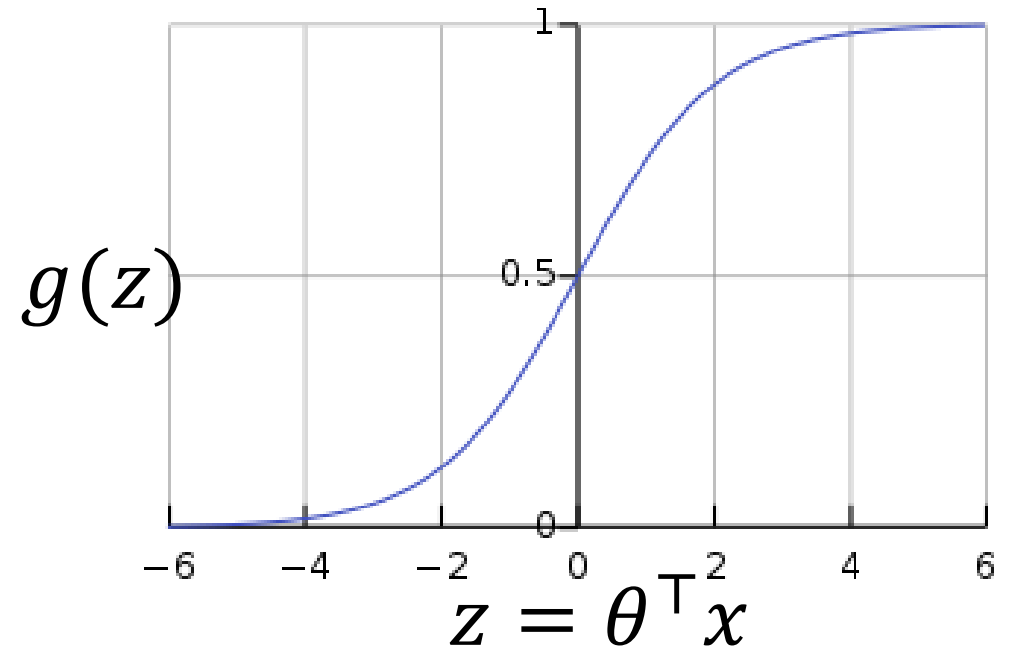
- $h_{\theta}(x)$ = estimated *probability* that $y = 1$ on input x
- *Example:* If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$
- $h_{\theta}(x) = 0.7$
- Tell patient that 70% chance of tumor being malignant

Logistic Regression

$$h_{\theta}(x) = g(\theta^{\top}x)$$

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

Log-linear model



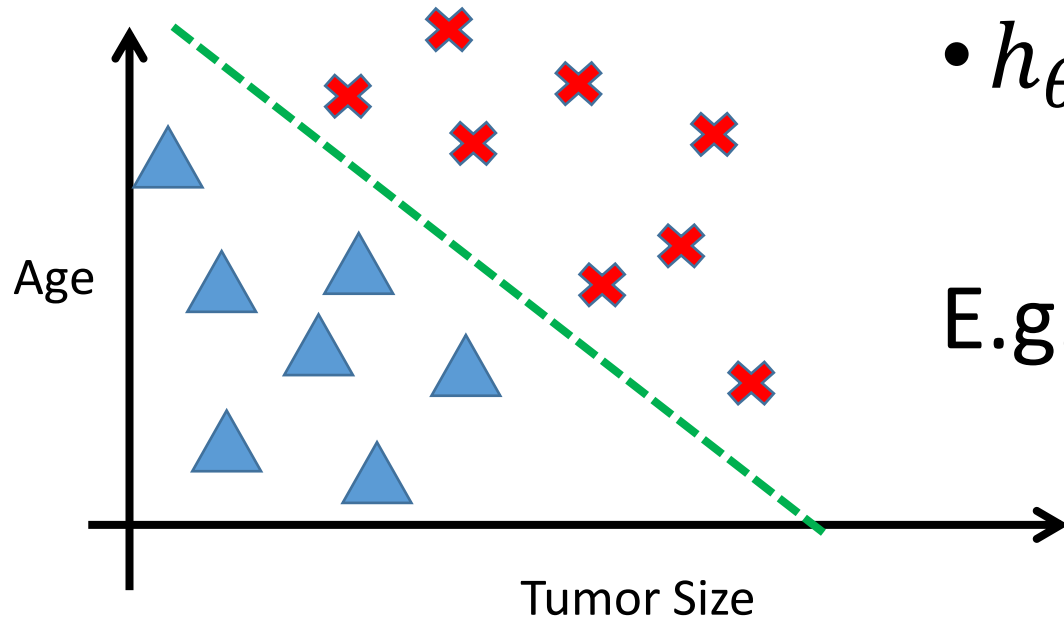
Suppose we predict “ $y = 1$ ” if $h_{\theta}(x) \geq 0.5$

$$z = \theta^{\top}x \geq 0$$

we predict “ $y = 0$ ” if $h_{\theta}(x) < 0.5$

$$z = \theta^{\top}x < 0$$

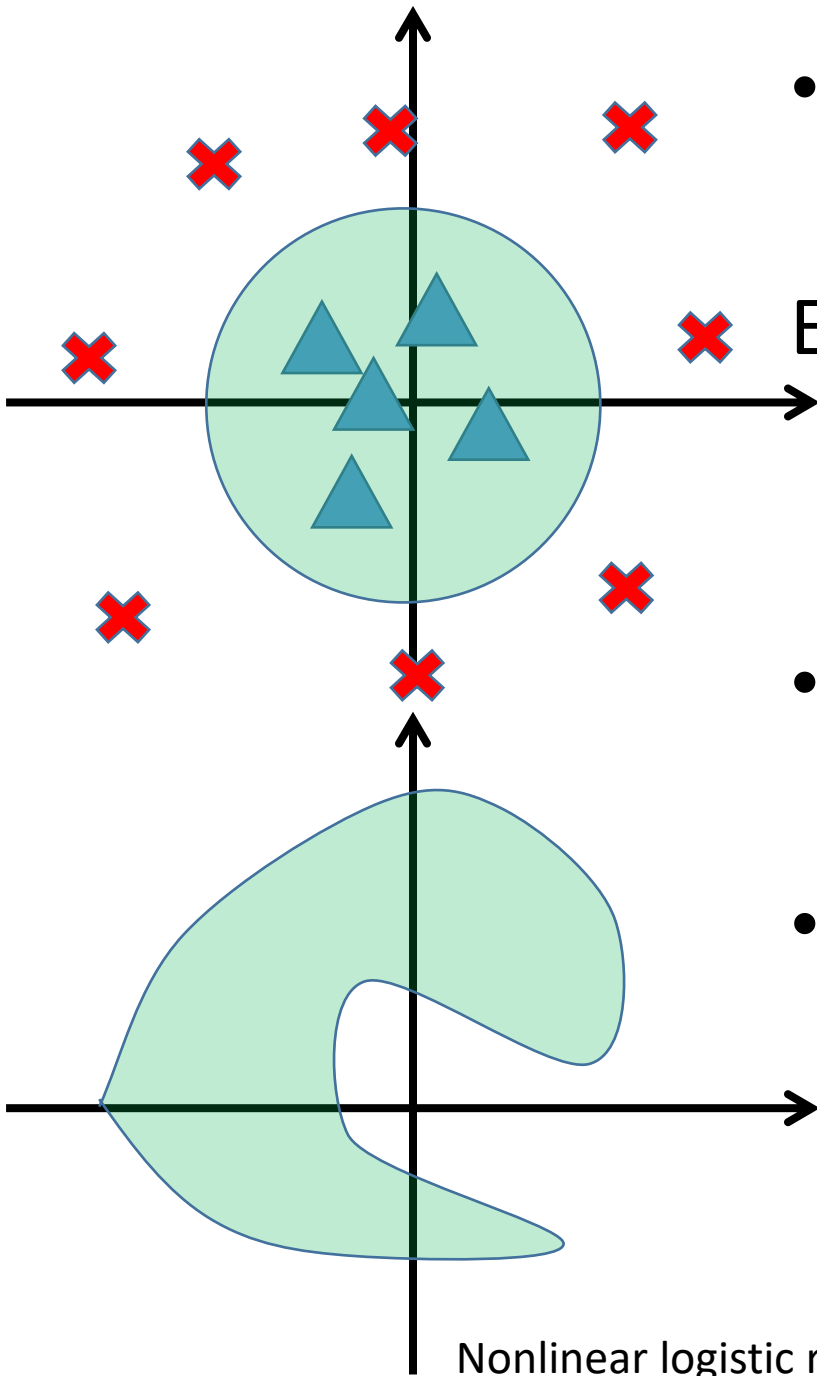
Decision Boundary



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

E.g., $\theta_0 = -3$, $\theta_1 = 1$, $\theta_2 = 1$

- Predict “ $y = 1$ ” if $-3 + x_1 + x_2 \geq 0$



- $h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$

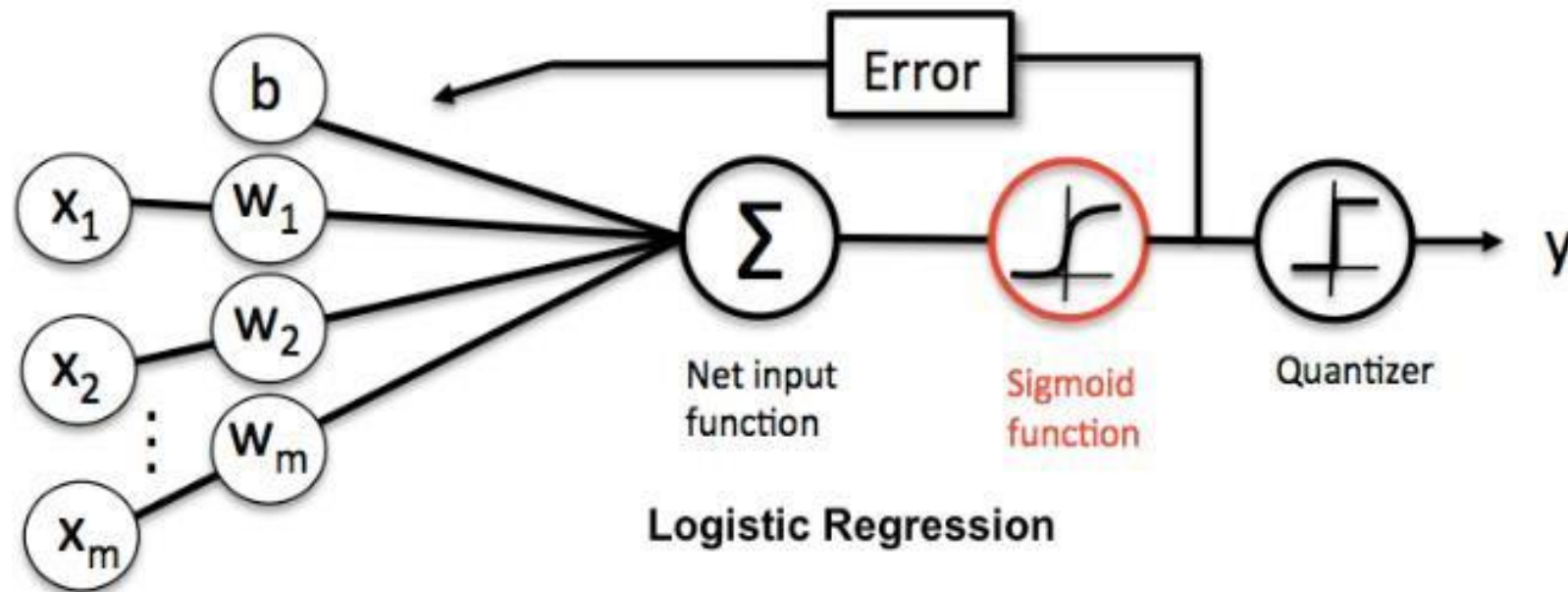
- E.g., $\theta_0 = -1, \theta_1 = 0, \theta_2 = 0, \theta_3 = 1, \theta_4 = 1$

- Predict “ $y = 1$ ” if $-1 + x_1^2 + x_2^2 \geq 0$

- $h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \dots)$

A multiplicative feature basis function!

Logistic Regressor Architecture



(Representation!)

Logistic Regression

- Hypothesis *representation*
- **Cost function (*evaluation*)**
- Logistic regression with gradient descent (*optimization*)
- Regularization
- Multi-class classification

Training set with m examples

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$x_0 = 1, y \in \{0, 1\}$$

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

How to choose parameters θ ?

Cost Function for Linear Regression

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)})$$

$$\text{Cost}(h_{\theta}(x), y) = \frac{1}{2} (h_{\theta}(x) - y)^2$$

Cost Function for Linear Regression

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y)$$

$$\text{Cost}(h_{\theta}(x), y) = \frac{1}{2} (h_{\theta}(x) - y)^2$$

Where does this cost come from?
Remember the Gaussian distribution?



Questions?

Deep robots!

Deep questions?!

