MACHINE LEARNING

LESSON 03:

Logistic Regression

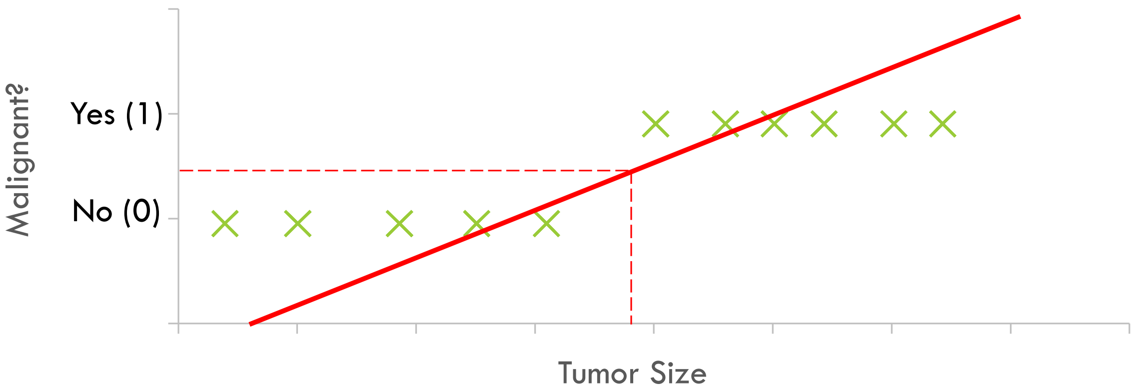
Inspired by and adapted from the Coursera Machine Learning course by Andrew Ng
(coursera.org/learn/machine-learning)

CLASSIFICATION

- Examples:
  - Email: Spam vs Not Spam
  - Tumor: Malignant vs Benign
  - Online Transactions Fraud: Yes vs No

- Binary (2-class) classification $\Rightarrow y \in \{0,1\}$
  - 0: negative class
  - 1: positive class

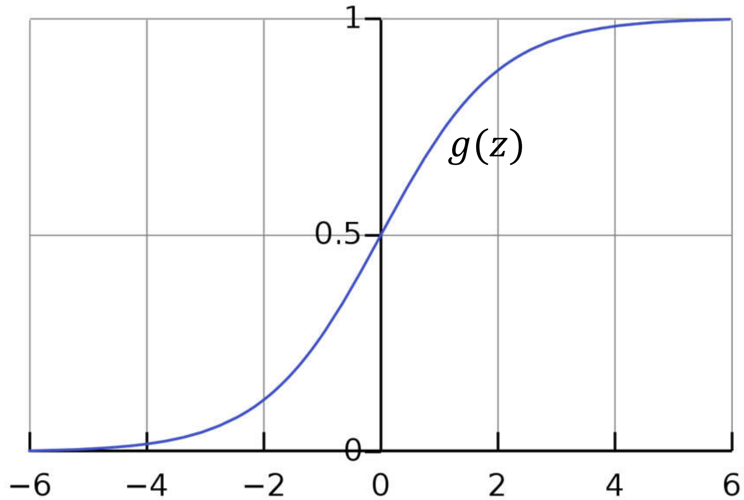
CLASSIFICATION



- $h(x)$ should have a **threshold** at 0.5
  - if $h(x) \geq 0.5 \Rightarrow$ predict that $y = 1$
  - if $h(x) < 0.5 \Rightarrow$ predict that $y = 0$

HYPOTHESIS REPRESENTATION

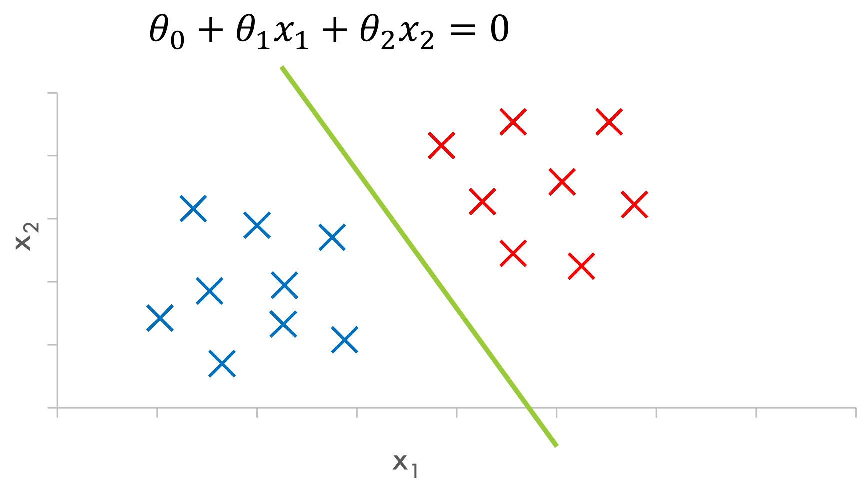
- We want $0 \leq h(x) \leq 1 \Rightarrow$ we add a function g
  - The hypothesis now becomes: $h(x) = g(\sum_{j=0}^{n} \theta_j x_j)$
  - Where $g(z) = \frac{1}{1+e^{-z}}$ (logistic/sigmoid function)



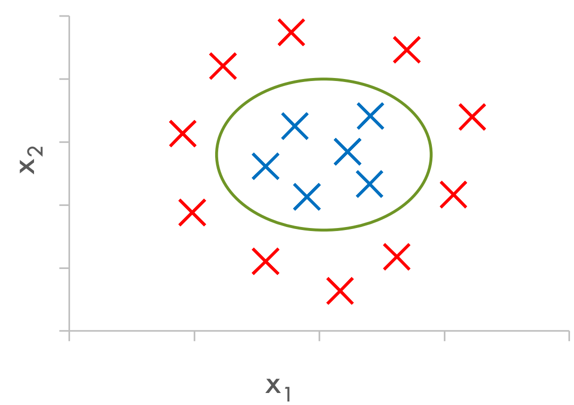
Interpretation of hypothesis output:
$h(x)$= estimated probability that y=1 given an input x

DECISION BOUNDARY

- Example:
  - $h(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$
  - We predict $y = 1$ if $h(x) \geq 0.5$
  $\Rightarrow \theta_0 + \theta_1 x_1 + \theta_2 x_2 \geq 0$



- More complex boundary
  - $h(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$

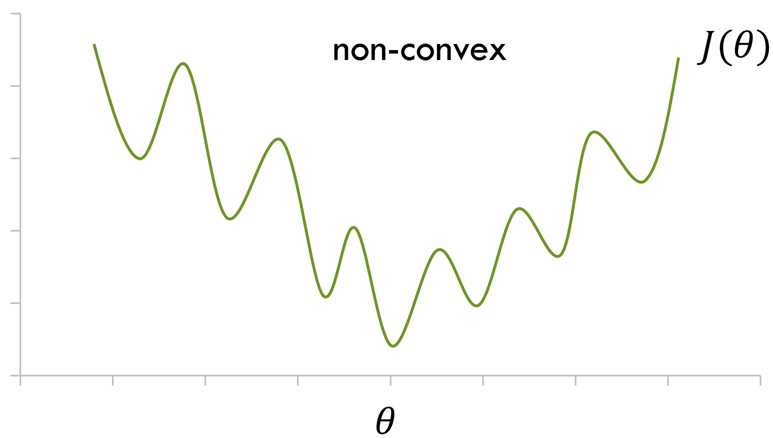


COST FUNCTION

- Again from linear regression
  - Hypothesis: $h(x) = \sum_{j=0}^{n} \theta_j x_j$ where $x_0 = 1$
  - Cost function: $J(\theta_0, \theta_1, \dots, \theta_n) = J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$

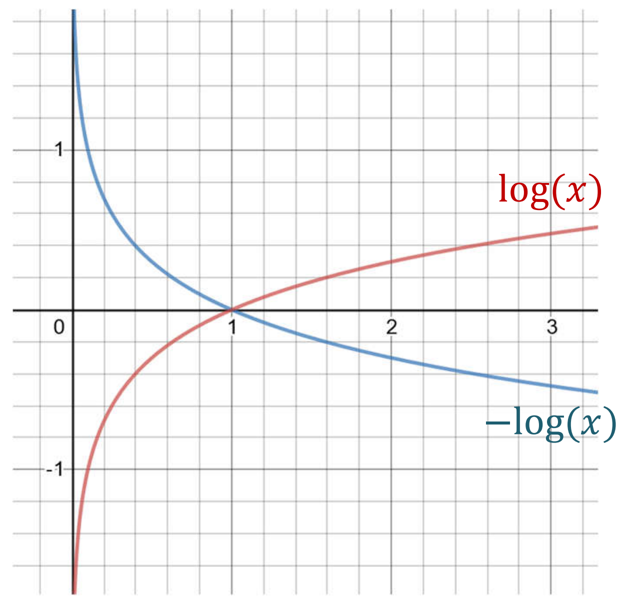
    where $y^{(i)}$ is either 0 or 1

    and $0 \leq h\left(x^{(i)}\right) \leq 1$



LOGISTIC REGRESSION COST FUNCTION

- $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} C\left(h\left(x^{(i)}\right), y^{(i)}\right) = \frac{1}{m}\sum_{i=1}^{m} C^{(i)}$
  - where $C^{(i)} = -y^{(i)} \log\left(h\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log(1 - h\left(x^{(i)}\right)$
  - $\Rightarrow C^{(i)} = \begin{cases} -\log\left(h\left(x^{(i)}\right)\right) & \text{if } y = 1 \\ -\log\left(1 - h\left(x^{(i)}\right)\right) & \text{if } y = 0 \end{cases}$



EXERCISE 3.1

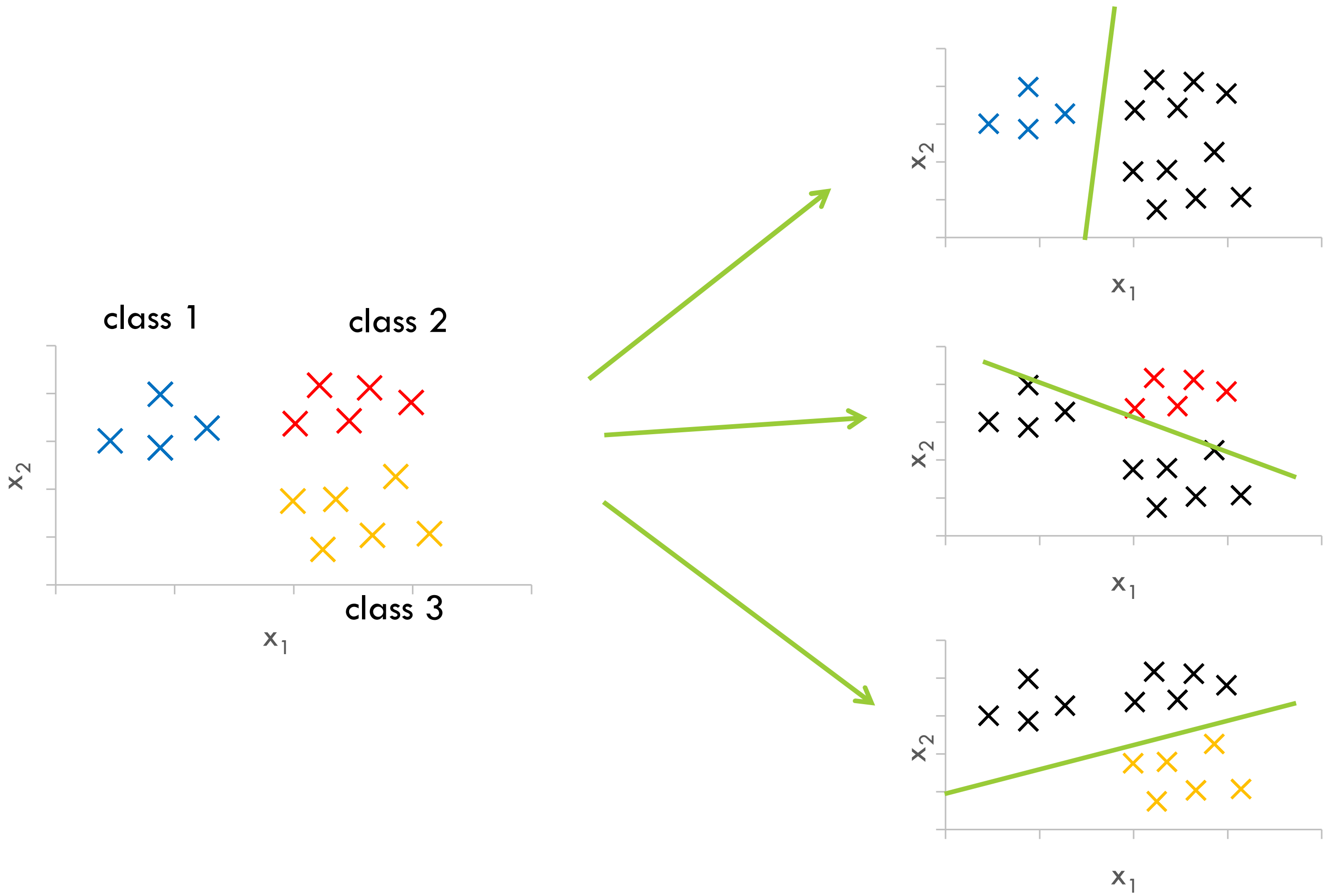
1. Given the dataset 'data_3_1_1.csv', build a logistic regression model with the hypothesis $h(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$
2. Given the dataset 'data_3_1_2.csv', build two logistic regression models:
   - $h_1(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$
   - $h_2(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$

   Compare these two models. Which one is better?

MULTICLASS CLASSIFICATION

- Example:
  - Emotion detection: Sad, Happy, Angry
  - Animal image recognition: Dog, Cat, Mouse, Bird, Fish
  - Weather forecast: Sunny, Cloudy, Rain, Snow

MULTICLASS CLASSIFICATION



$h_1(x) = 0.8$

$h_2(x) = 0.1$

$h_3(x) = 0.5$

EXERCISE 3.2

Given the dataset 'data_3_2.csv', build a multiclass (4-class) classification model to fit this data. You can choose your own hypothesis.