

Machine Learning, Spring 2019

Regression

Reading Assignment: Chapter 5 & 6

Python tutorial: <http://learnpython.org/>

TensorFlow tutorial: <https://www.tensorflow.org/tutorials/>

PyTorch tutorial: <https://pytorch.org/tutorials/>

Announcement

- 20% bonus points on Mini Project One

Polynomial Linear Regression

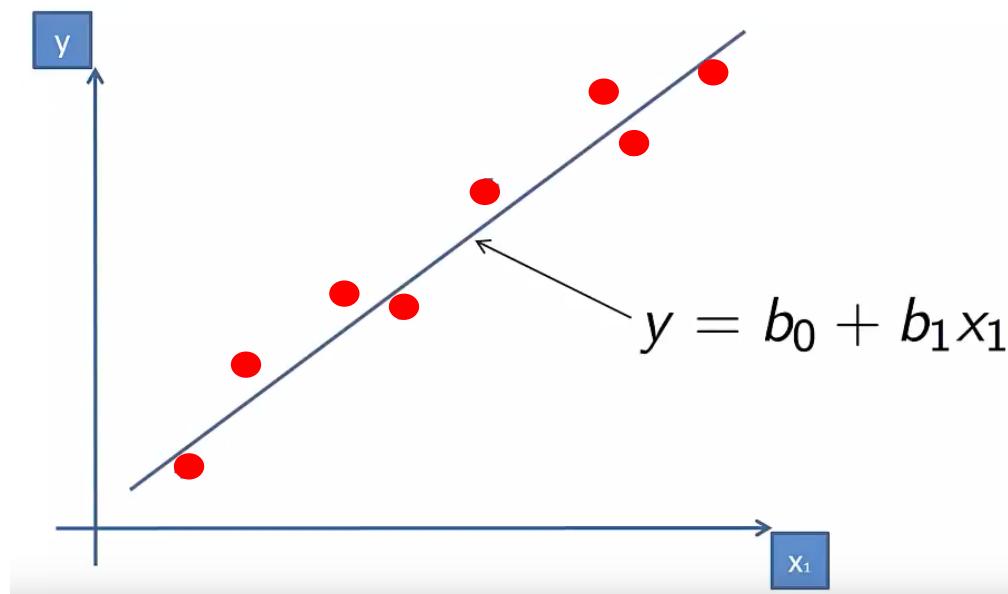
Linear Regression:

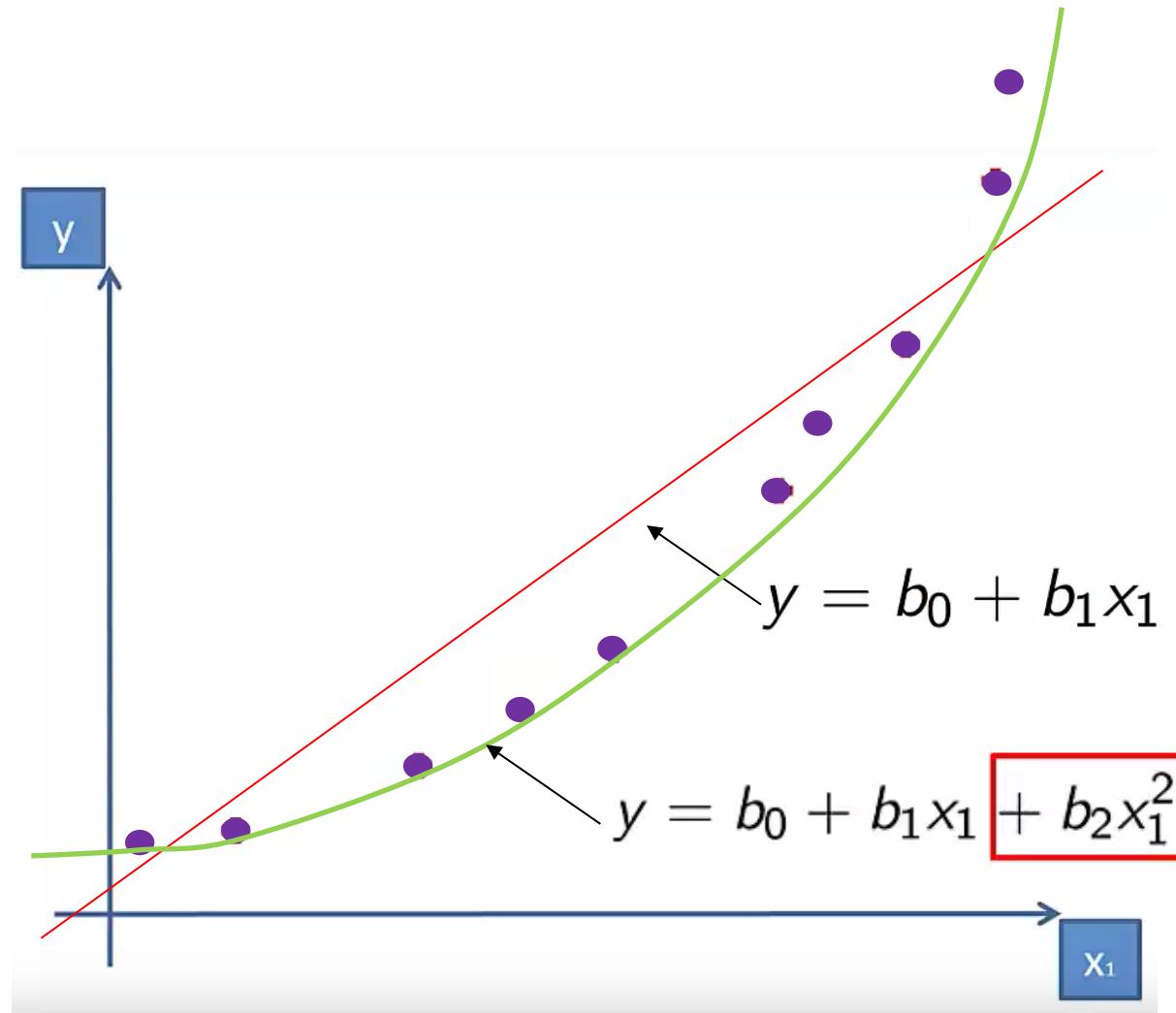
$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Polynomial Linear Regression:

$$y = b_0 + b_1x_1 + b_2x_1^2 + \dots + b_nx_1^n$$

Why polynomial regression





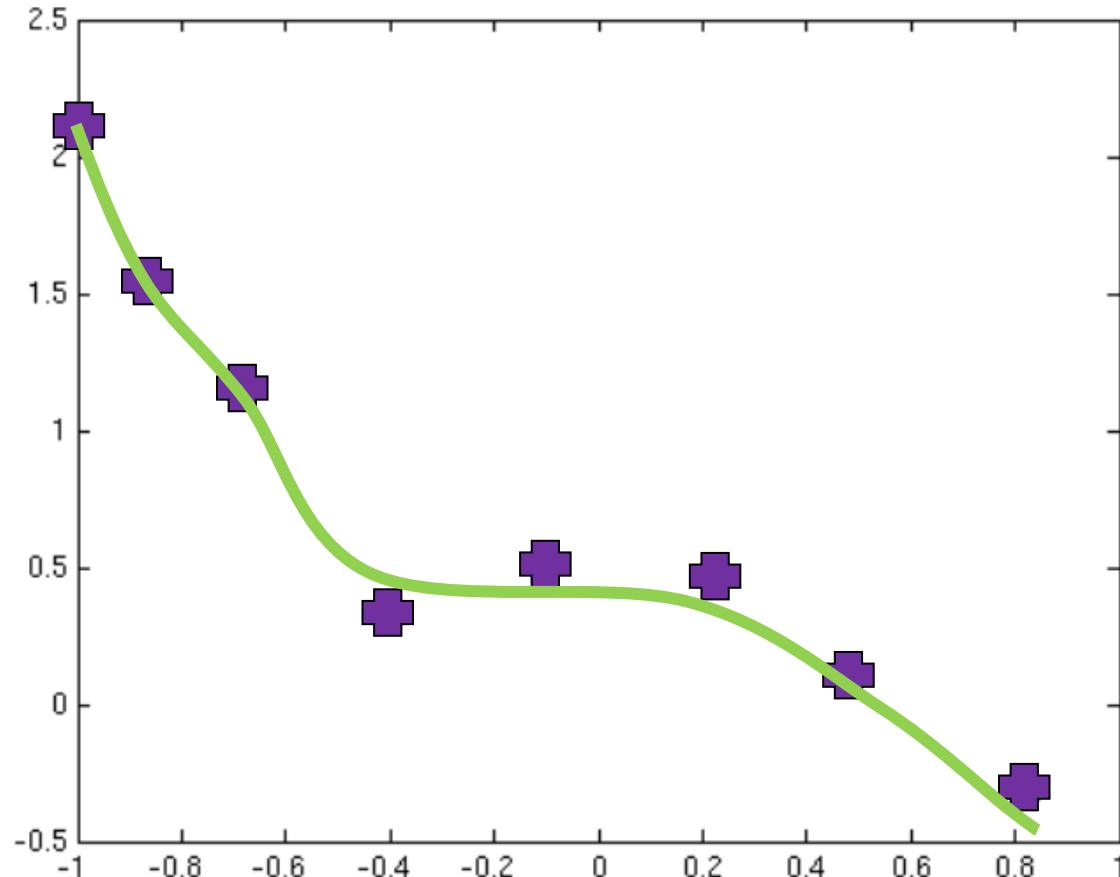
Why still called “Linear”?

Polynomial Linear Regression:

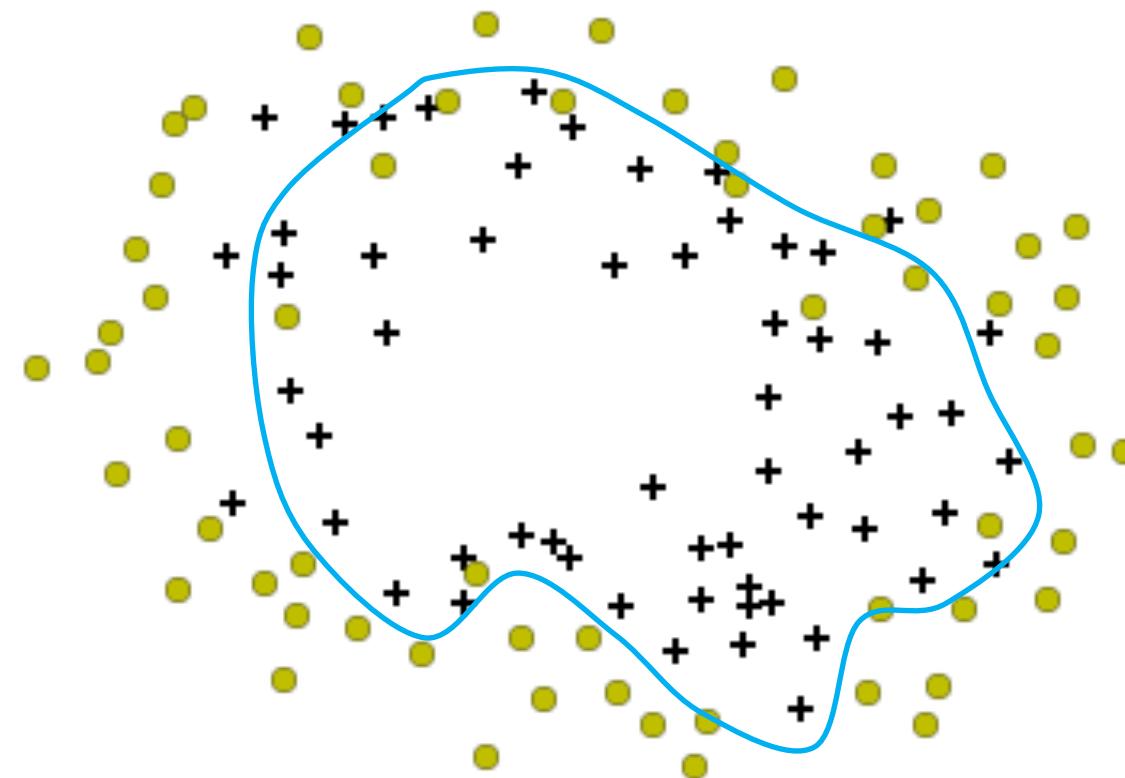
$$y = b_0 + b_1x_1 + b_2x_1^2 + \dots + b_nx_1^n$$

The function is expressed as the linear combination of unknowns (i.e. coefficients)?

Polynomial Regression

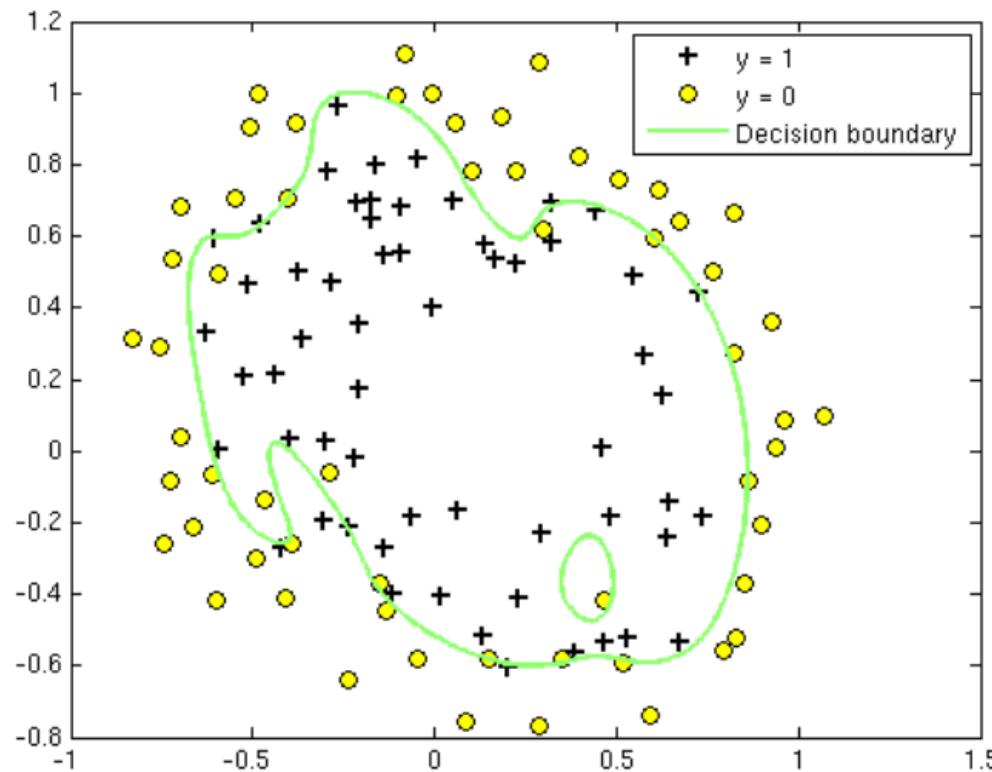


$$Y = \theta^T \mathbf{x}$$



$$Y = \sigma(\theta^T \mathbf{x})$$

Over Fitting



Regularized logistic regression

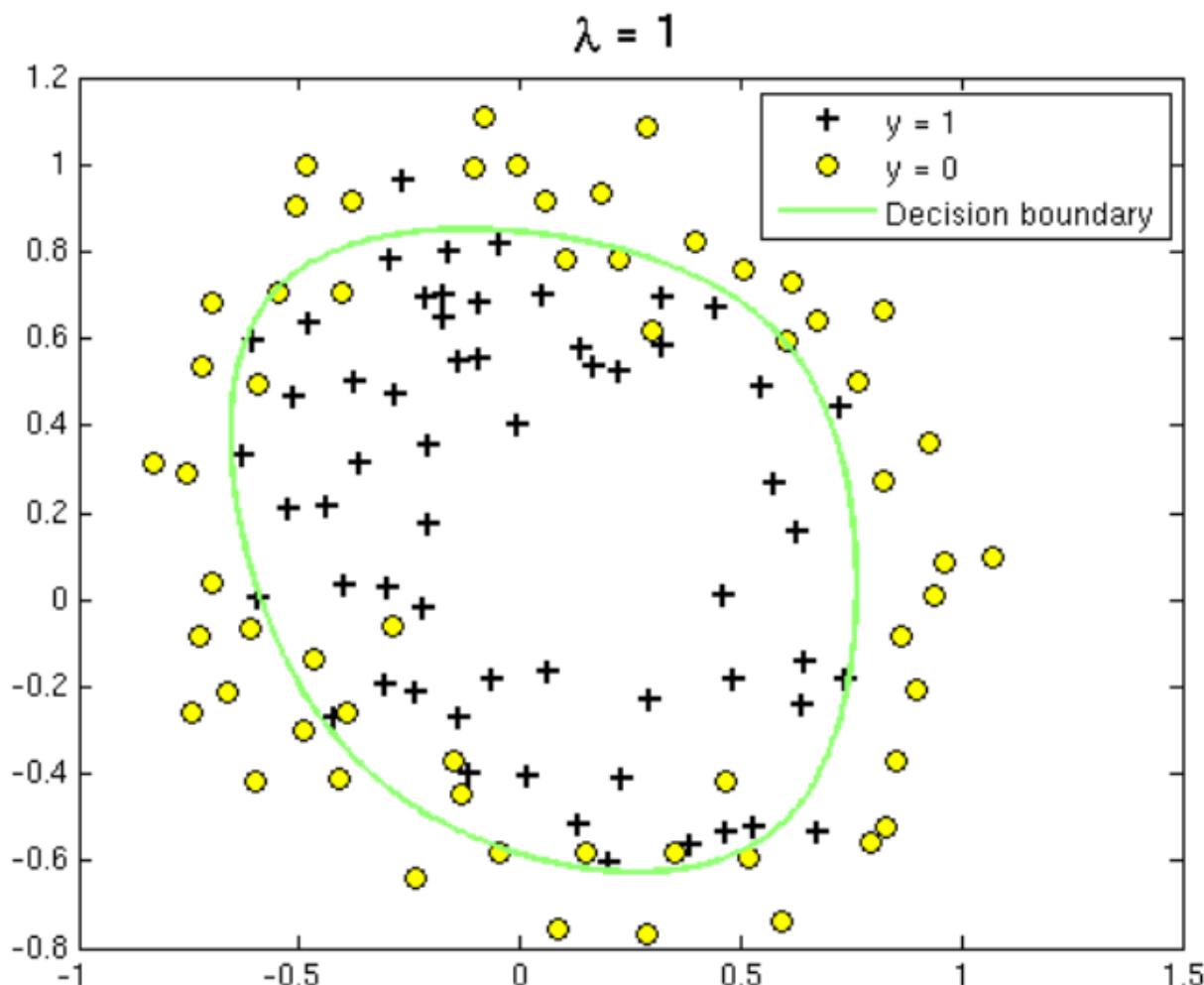
Regularized logistic regression Loss

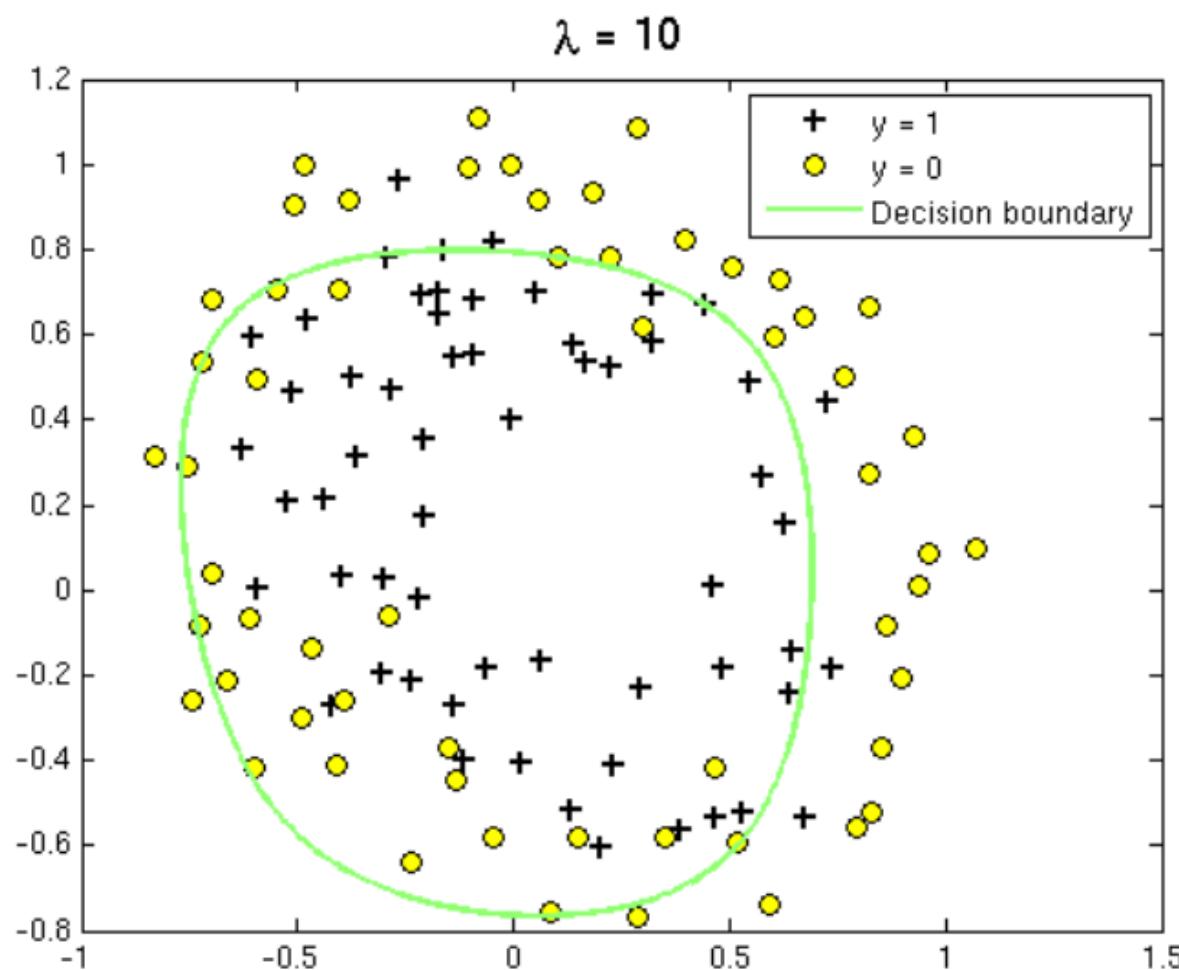
Loss function:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

Gradient function:

$$\begin{aligned} \frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \left[-\frac{1}{m} \sum_{i=1}^m \left(y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right) + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2 \right] \\ &= \frac{1}{m} \sum_{i=1}^m \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} + \frac{\lambda}{m} \theta_j \end{aligned}$$





Machine Learning, Spring 2019

Classification

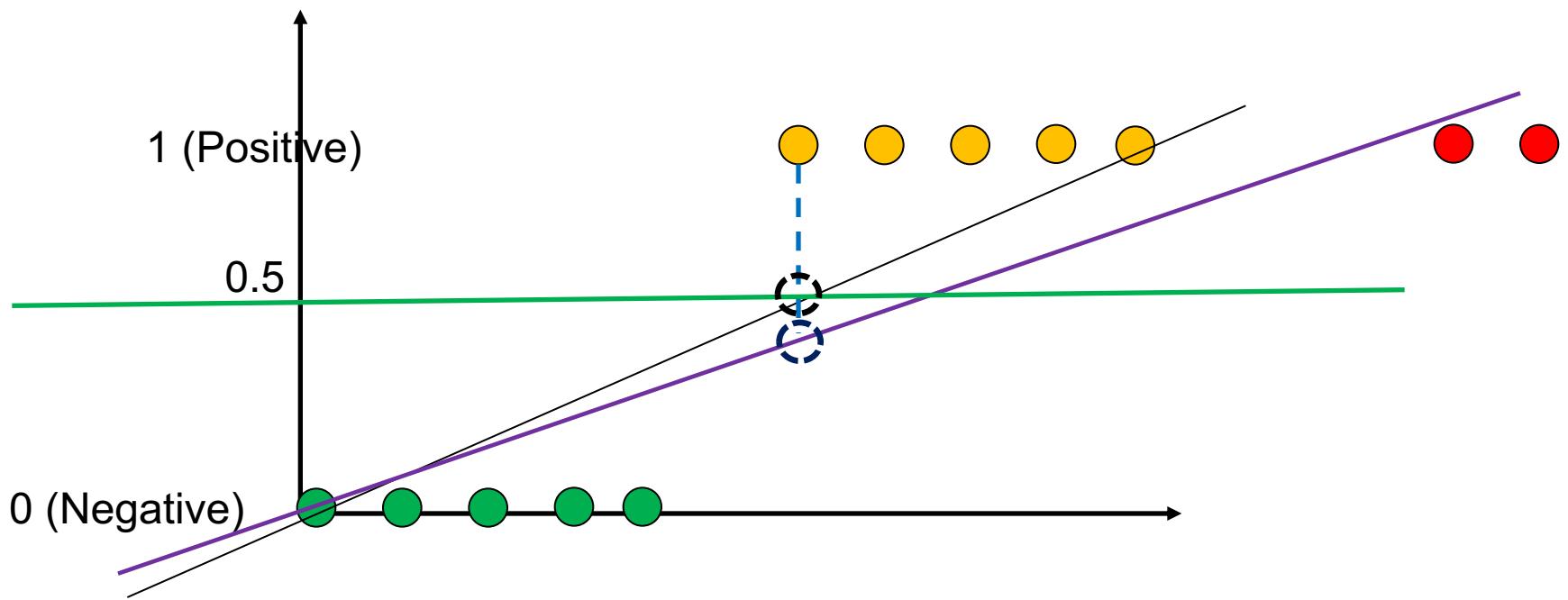
Overview

Problems:

1. Spam or Non-spam emails
2. Person re-identification (same or different person)



Linear Regression or Logistic Regression?



Hypothesis

- Hypothesis: A hypothesis is a certain function that we believe (or hope) is similar to the true function, the target function that we want to model. In context of email spam classification, it would be the rule we came up with that allows us to separate spam from non-spam emails.

Classifier

- Classifier: A classifier is a special case of a hypothesis (nowadays, often learned by a machine learning algorithm). A classifier is a hypothesis or discrete-valued function that is used to assign (categorical) class labels to particular data points. In the email classification example, this classifier could be a hypothesis for labeling emails as spam or non-spam. However, a hypothesis must not necessarily be synonymous to a classifier. In a different application, our hypothesis could be a function for mapping study time and educational backgrounds of students to their future SAT scores.

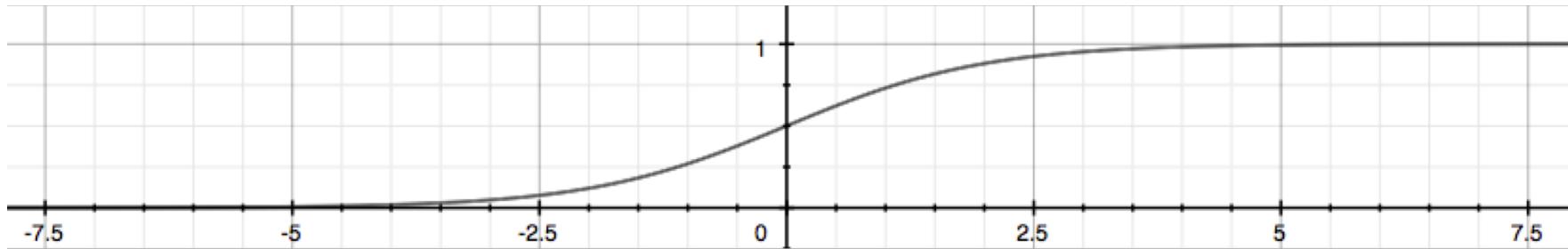
Logistic Regression Hypothesis

$$h_{\theta}(x) = g(\theta^T x) \quad (1)$$

$$g(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

$$z = \theta^T x \quad (3)$$

$$g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (4)$$



Hypothesis Function

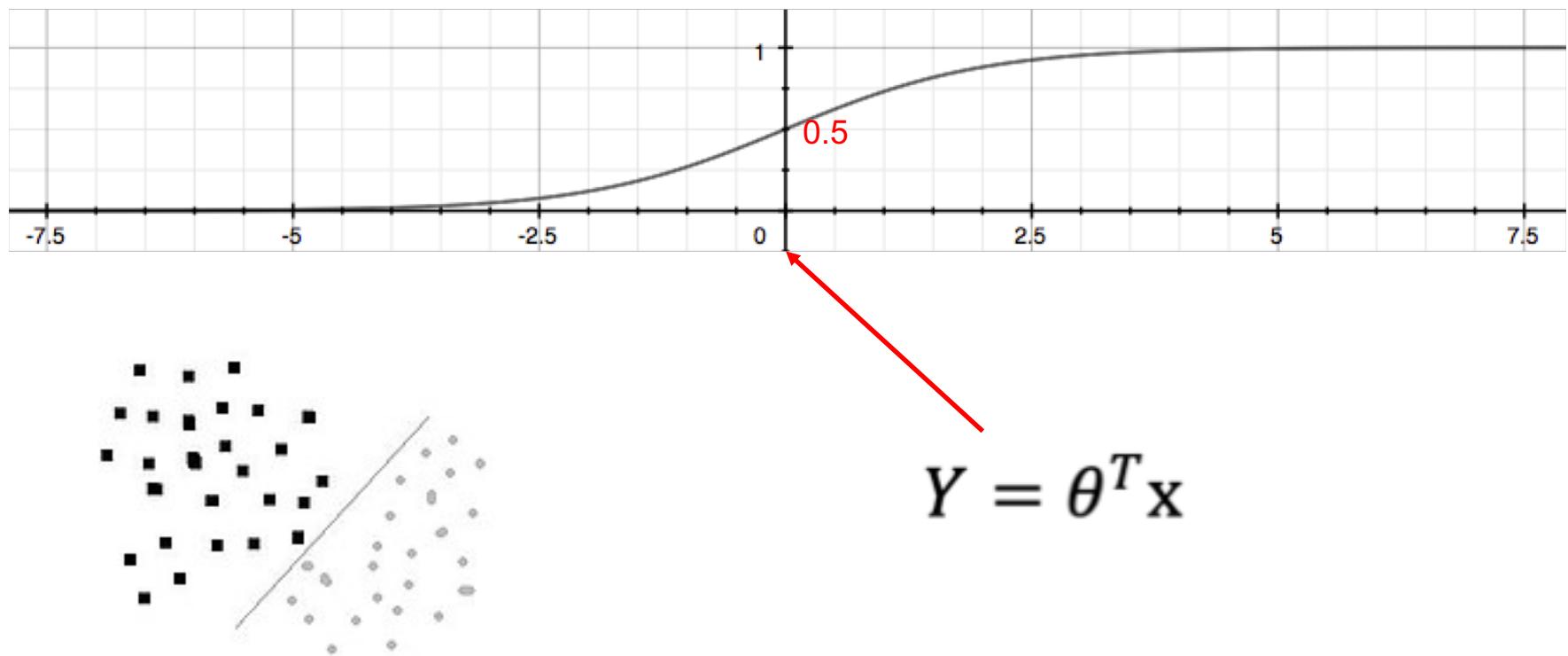
We have: $h_{\theta}(x) = P(y = 1|x; \theta)$

$$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$$

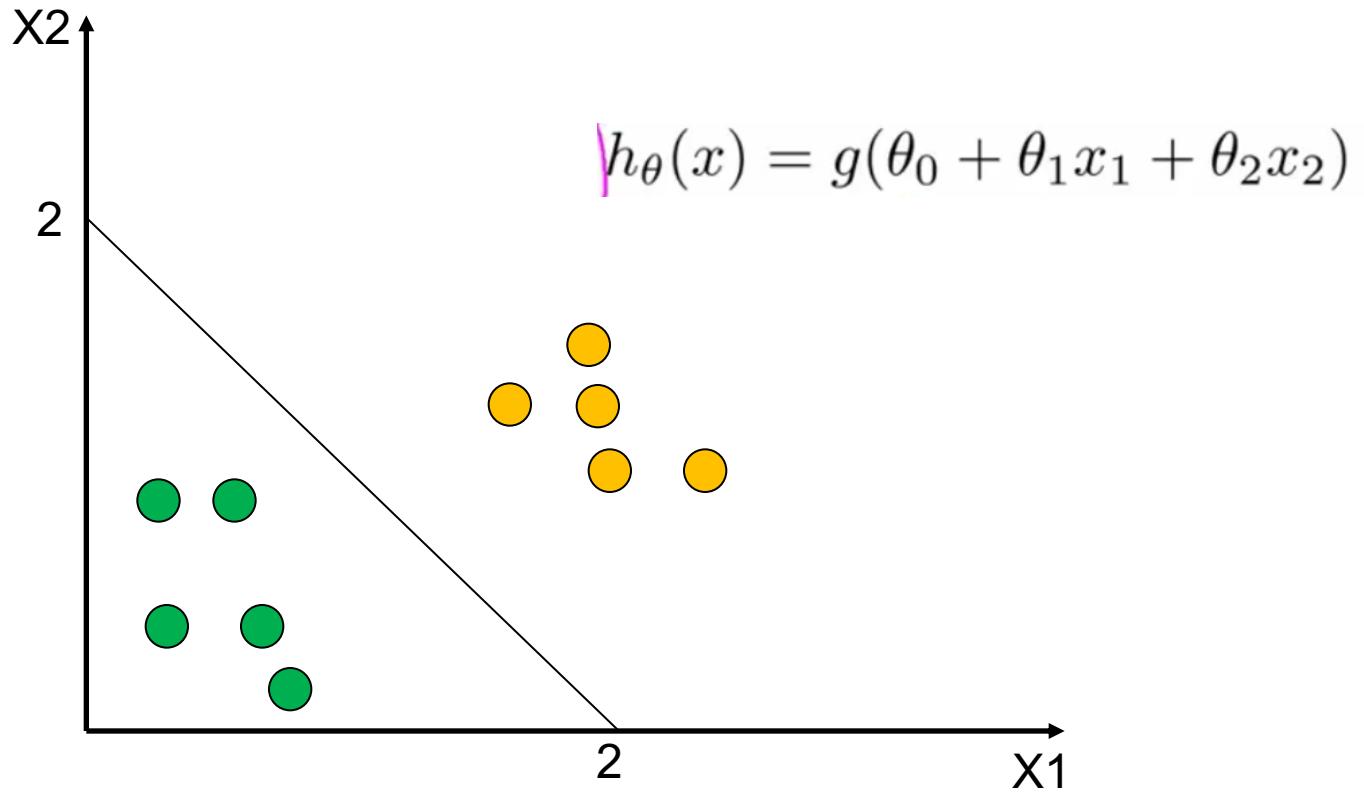
Then:

$$h_{\theta}(x) = P(y = 1|x; \theta) = 1 - P(y = 0|x; \theta)$$

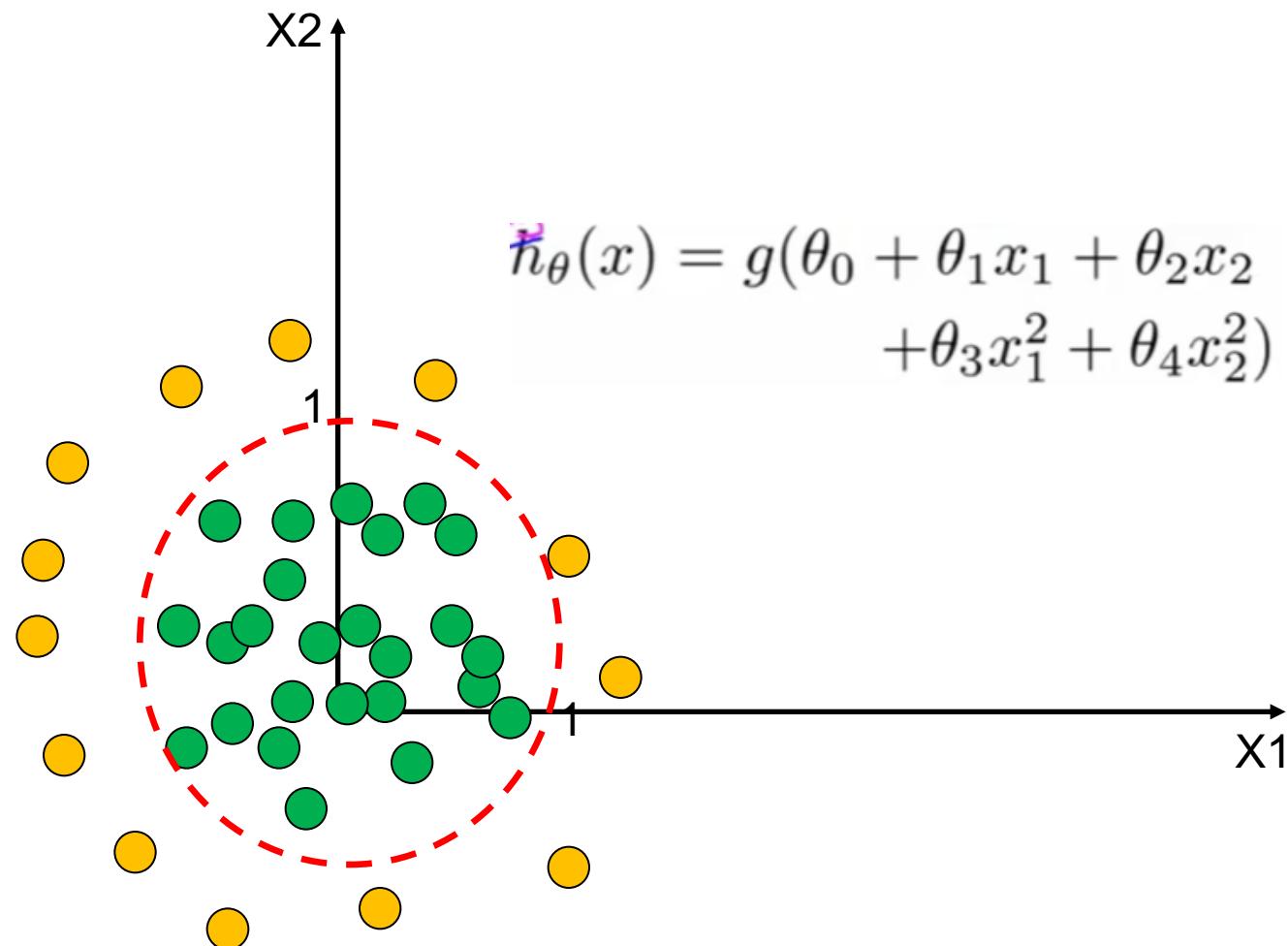
Decision Boundary



Classification: Linearly Separable



Classification: Non-Linearly Separable



Machine Learning, Spring 2019

Classification Cost Function

Design of Cost Function

Training Samples:

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

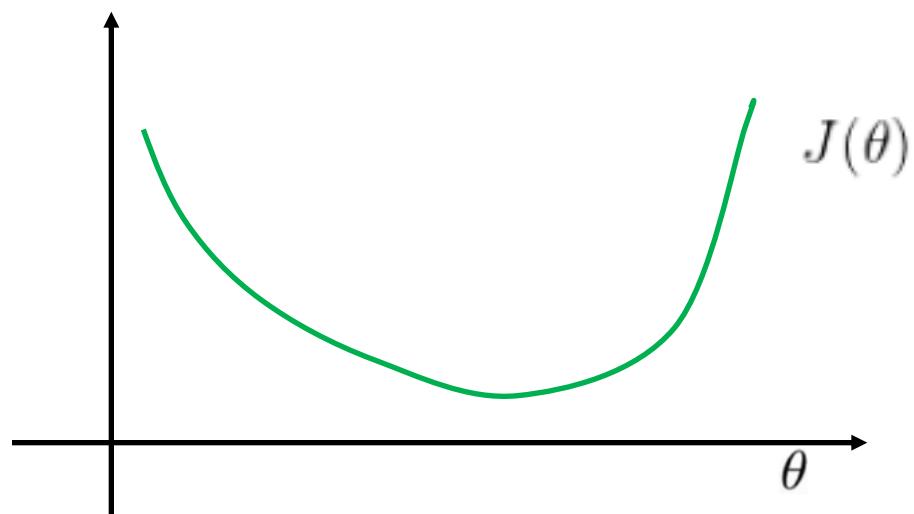
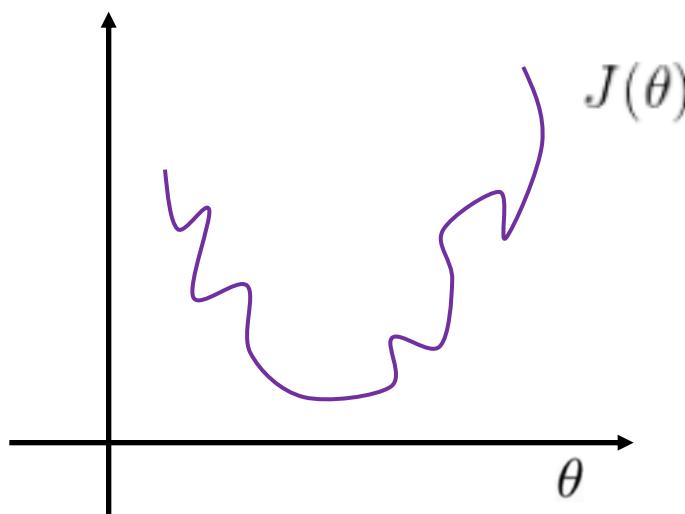
Hypothesis:

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

Direction Difference function:

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

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



Log Likelihood function:

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

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

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

We can write it as one function as:

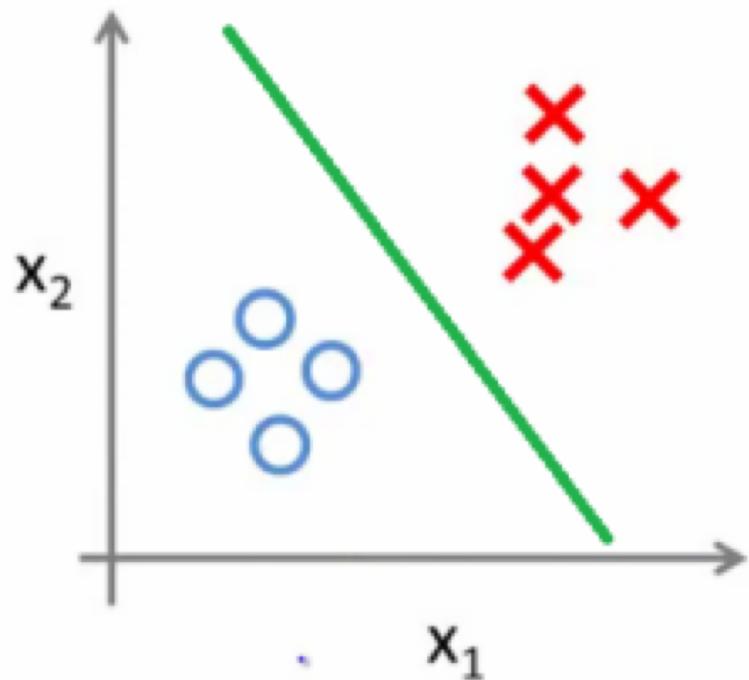
$$cost(h_\theta(x), y) = -y \log(h_\theta(x)) + [-(1 - y) \log(1 - h_\theta(x))]$$

Machine Learning, Spring 2019

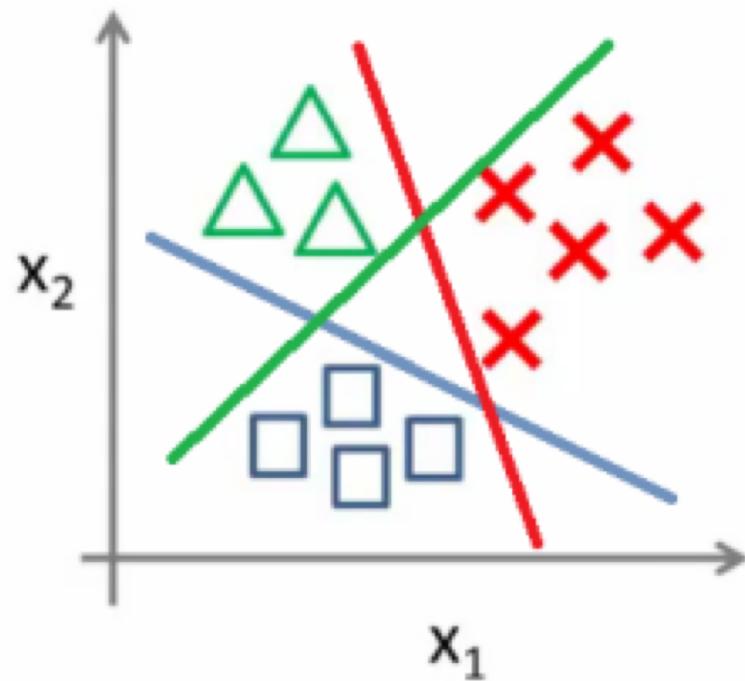
One-vs-All Classification

Multi-class classification

Binary classification:



Multi-class classification:



Hypothesis

Logistic Regression Model:

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

Log Likelihood function:

Cost Function

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

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

$$y \in \{0, 1, \dots, n\}$$

$$h_\theta^{(0)}(x) = P(y = 0|x; \theta)$$

$$h_\theta^{(1)}(x) = P(y = 1|x; \theta)$$

...

$$Cost(h_\theta(x), y) = -\log(h_\theta(x))$$

$$h_\theta^{(n)}(x) = P(y = n|x; \theta)$$