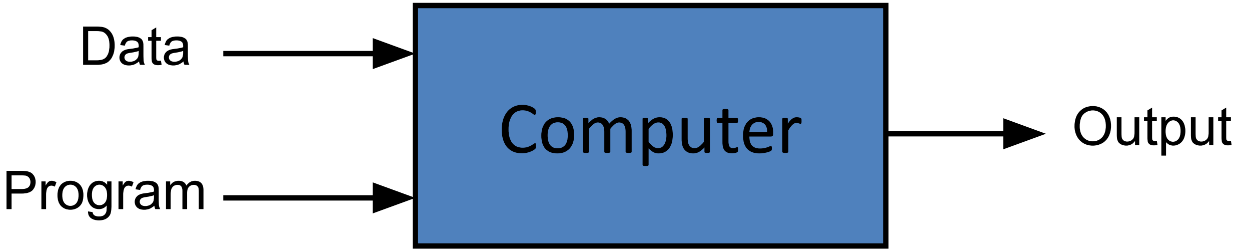
Introduction to **Machine Learning**
&
Linear Regression – **Model Representation**

Acknowledgments: Andrew Ng and Pedro Domingos

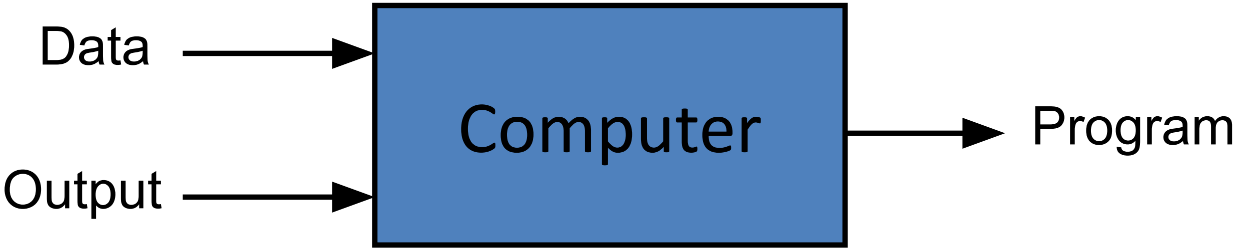
So What Is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Let the **data** do the work!

Traditional Programming



Machine Learning



Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining

  Large datasets from growth of automation/web.

  E.g., Web click data, medical records, biology, engineering
- Applications can’t program by hand.

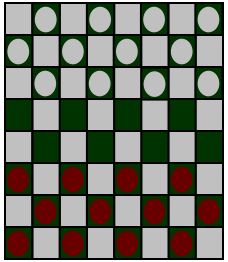
  E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.

Practical Applications of Machine Learning

- Spam filtering
- Speech/handwriting recognition
- Object detection/recognition
- Weather prediction
- Stock market analysis
- Search engines (e.g, Google)
- Ad placement on websites
- Credit-card fraud detection
- Webpage clustering (e.g., Google News)
- Social Network Analysis
- Recommendation systems (e.g., Netflix, Amazon)
- Automatic vehicle navigation

Machine Learning definition

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.



- Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

"A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

a) Watching you label emails as spam or not spam.

b) Classifying emails as spam or not spam.

c) The number (or fraction) of emails correctly classified as spam/not spam.

d) None of the above—this is not a machine learning problem.

Machine learning algorithms:

- Supervised learning
- Unsupervised learning

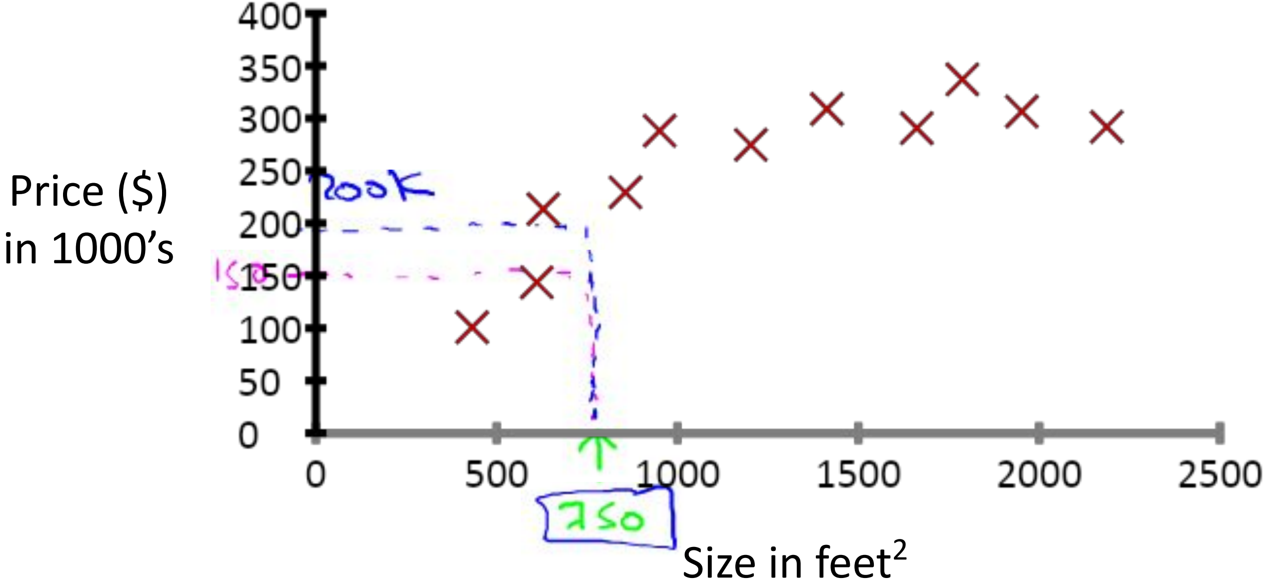
Others: recommender systems.

Introduction

to

Supervised Learning

Housing price prediction.

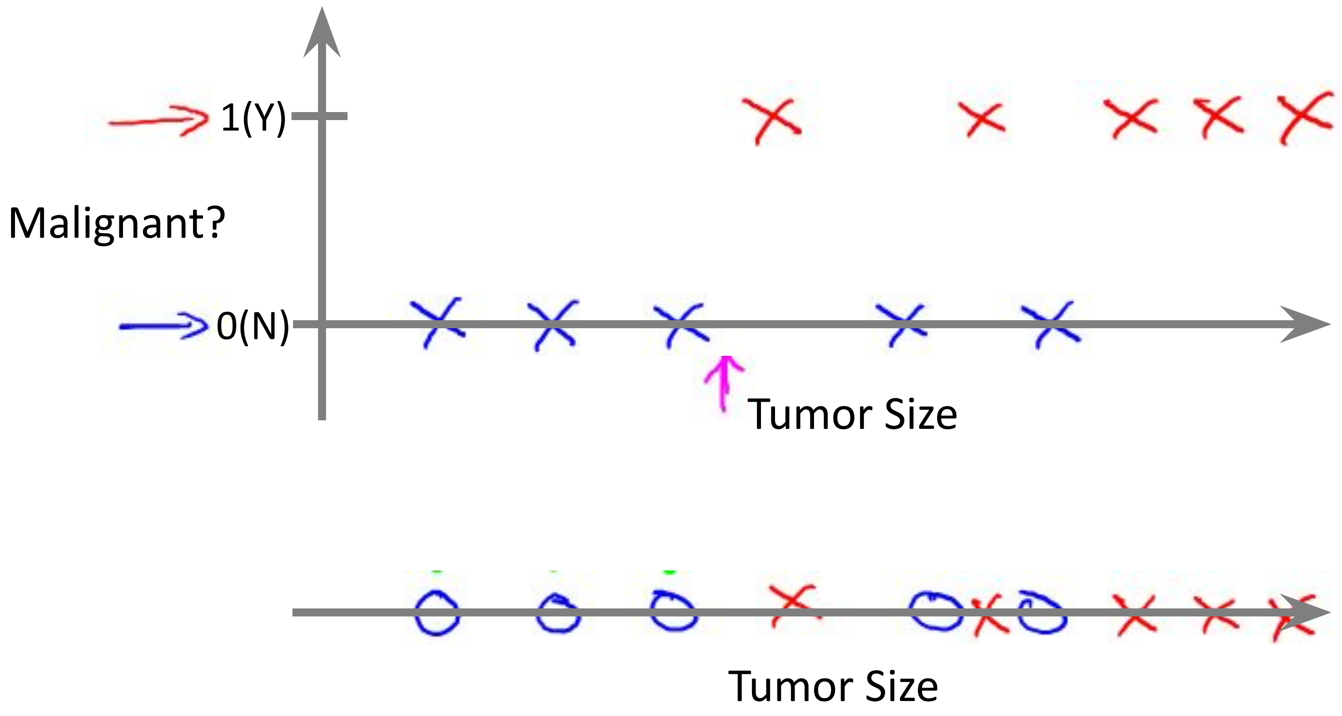


<u>Supervised Learning</u>

"right answers" given

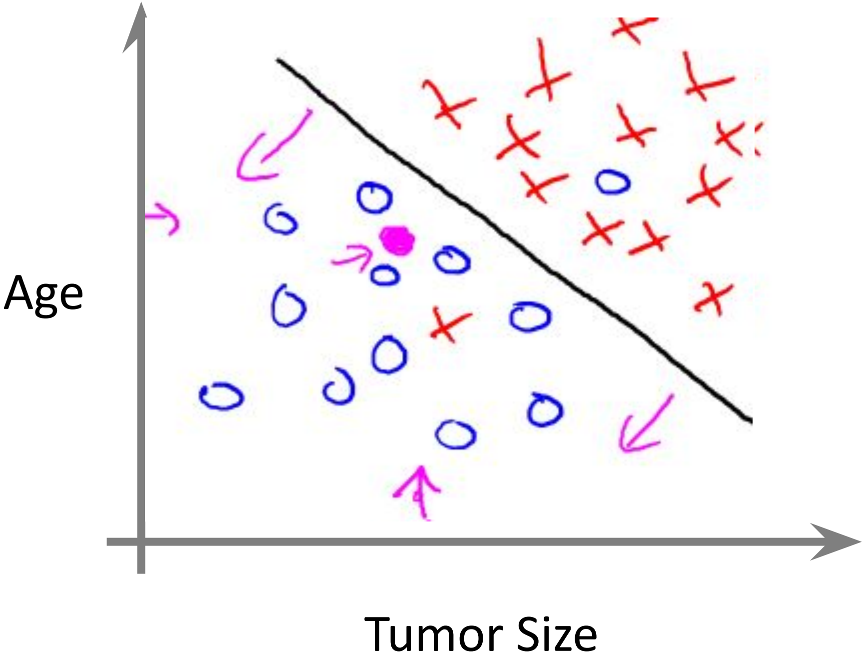
<u>Regression:</u> Predict continuous valued output (price)

Cancer (malignant, benign)



Classification

Discrete valued output (0 or 1)



- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

You’re running a company, and you want to develop learning algorithms to address each of two problems.

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
Problem 2: You’d like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

Should you treat these as classification or as regression problems?

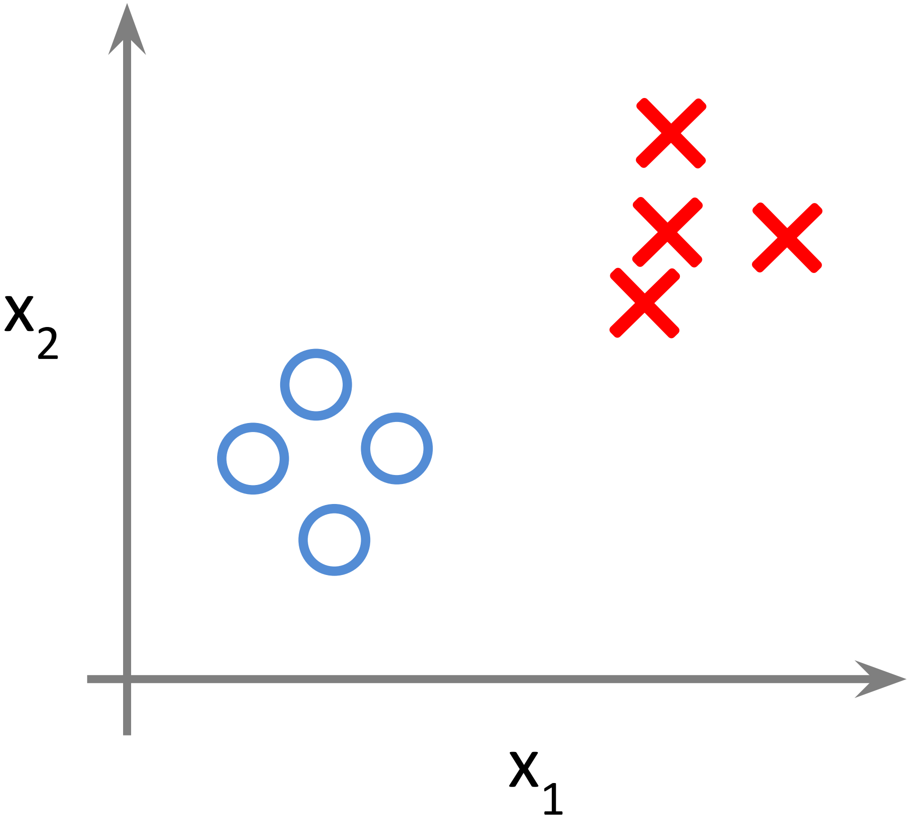
a) Treat both as classification problems.

b) Treat problem 1 as a classification problem, problem 2 as a regression problem.

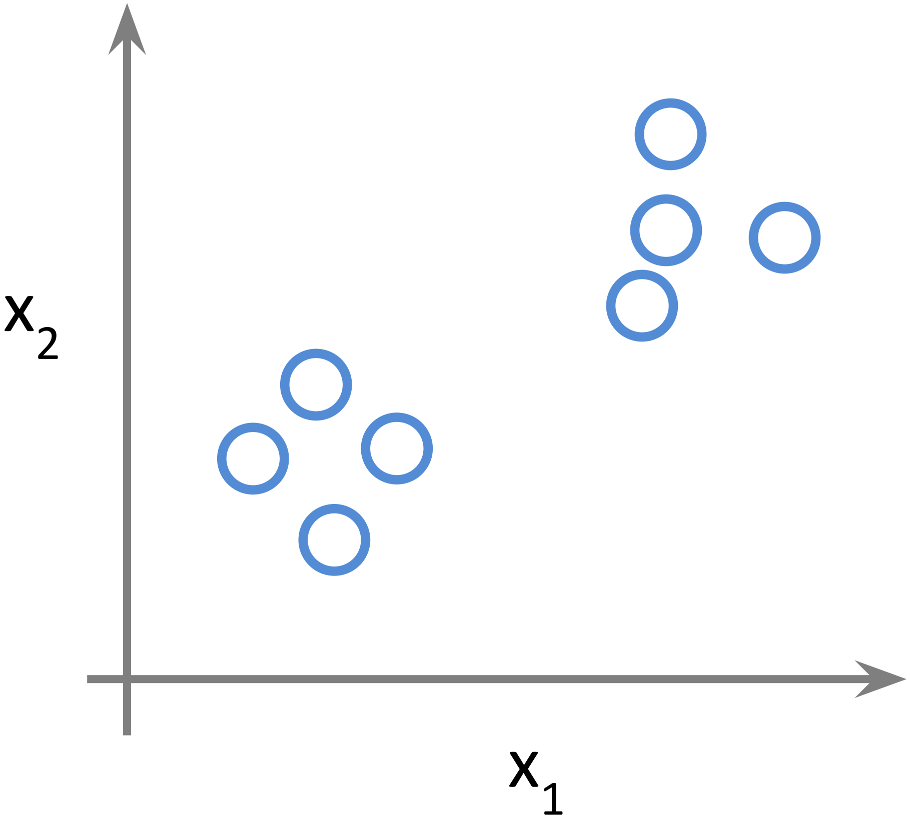
c) Treat problem 1 as a regression problem, problem 2 as a classification problem.
d) Treat both as regression problems.

Introduction to **Unsupervised Learning**

Supervised Learning



Unsupervised Learning



Google News

news.google.com

Web Images Videos Maps **News** Shopping Gmail more ▾ — Web History | Settings ▾ | Sign out

Google news

Search News | Search the Web

Advanced news search

U.S. edition ▾ Add a section »

Top Stories
- Deepwater Horizon
- Fed meeting
- Foreign exchange market
- Lindsay Lohan
- IBM
- Tom Brady
- Toronto International Film Festival
- Paris Hilton
- Iran
- Hurricane Igor

- Starred ☆
- San Francisco Bay Area
- World
- U.S.
- Business
- Sci/Tech
- More Top Stories
- Spotlight
- Health
- Sports
- Entertainment

All news
- Headlines
- Images

Top Stories

Christine O'Donnell »

White House official denies Tea Party-focused ad campaign

CNN Interna...

CNN International - Ed Henry - 1 hour ago

Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) -- A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...

Tea Party is misplacing the blame, former President Bill Clinton claims New York Daily News

GOP tea party backer defends Christine O'Donnell The Associated Press

Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon

all 726 news articles »

US Stocks Climb After Recession Called Over, Homebuilders Gain

MyFox Phila...

MarketWatch - Kristina Peterson - 16 minutes ago

NEW YORK (MarketWatch) -- US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.

Longest recession since 1930s ended in June 2009, group says Los Angeles Times

Downturn Was Longest in Decades, Panel Confirms New York Times

Wall Street Journal - AFP - CNN - USA Today

all 276 news articles »

Deepwater Horizon »

BP Oil Well, Site of National Catastrophe, Dies at One

Reuters

Vanity Fair - Juli Weiner - 22 minutes ago

The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.

+ Video: Blown-out BP Well Finally Killed in Gulf The Associated Press

Weiss Doubts BP Would End Operations in Gulf of Mexico: Video Bloomberg

CNN International - Wall Street Journal (blog) - The Guardian - New York Times

all 2,292 news articles »

Recent

Recession officially ended in June 2009
CNNMoney - Chris Isidore - 39 minutes ago

Hurricane Igor lashes Bermuda
USA Today - Gerry Broome - 5 minutes ago

'Explain what you want from us,' reads front-page editorial
msnbc.com - Olivia Torres - 10 minutes ago

Crisis response: Pakistan floods

San Francisco Bay Area - Edit

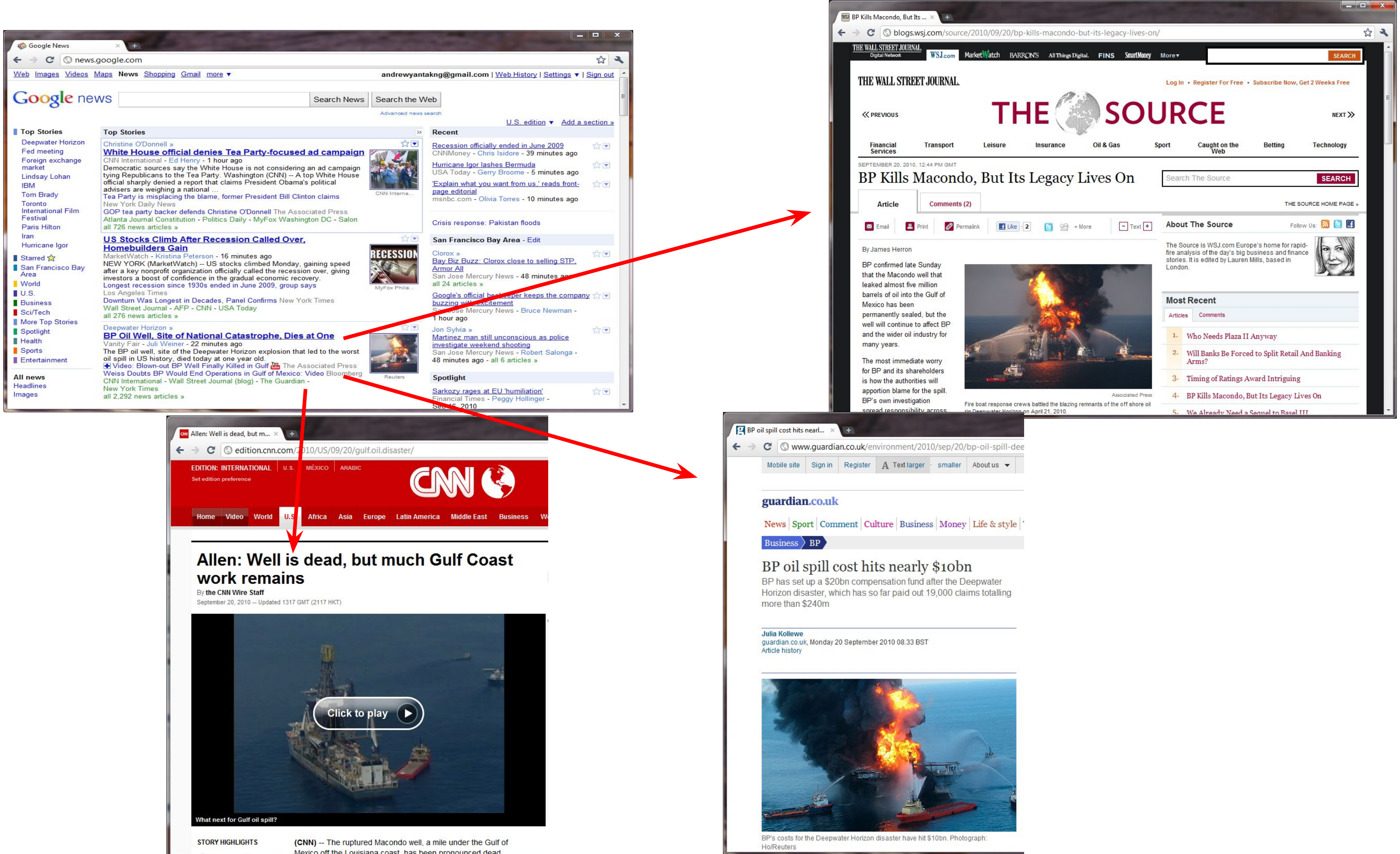
Clorox »
Bay Biz Buzz: Clorox close to selling STP, Armor All
San Jose Mercury News - 48 minutes ago - all 24 articles »

Google's official beekeeper keeps the company buzzing with excitement
San Jose Mercury News - Bruce Newman - 1 hour ago

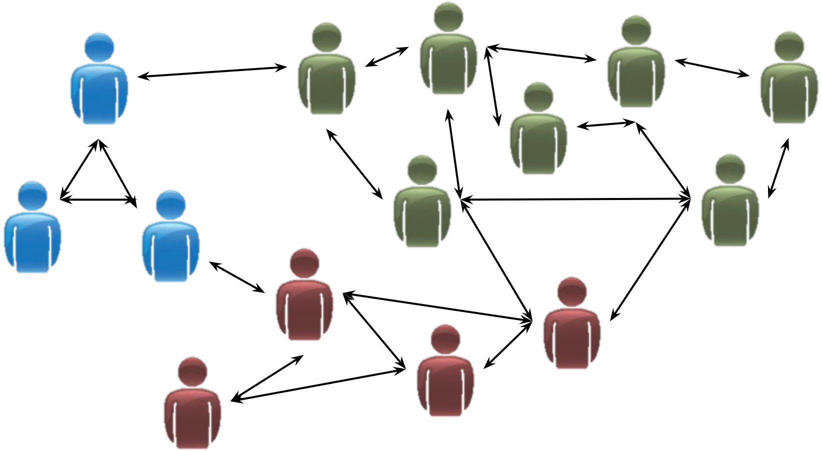
Jon Sylvia »
Martinez man still unconscious as police investigate weekend shooting
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Spotlight

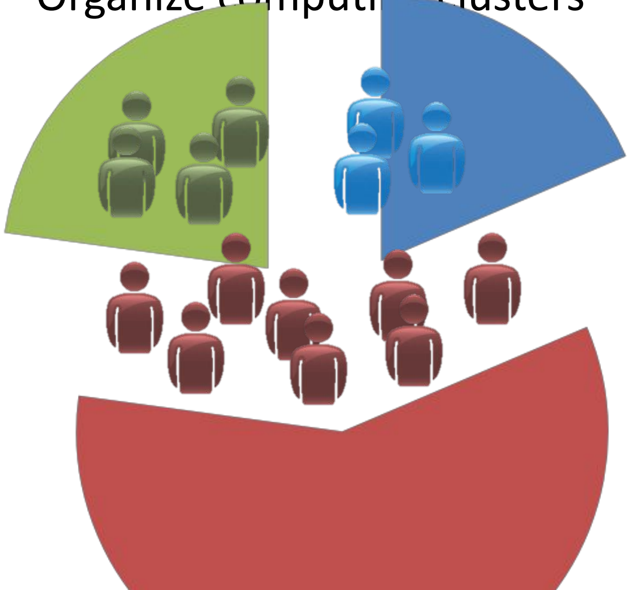
Sarkozy rages at EU 'humiliation'
Financial Times - Peggy Hollinger -


Google News
news.google.com
Top Stories
White House official denies Tea Party-focused ad campaign
US Stocks Climb After Recession Called Over, Homebuilders Gain
BP Oil Well, Site of National Catastrophe, Dies at One
Vanity Fair - Juli Weiner - 22 minutes ago
The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.
Recent
San Francisco Bay Area
Spotlight
BP Kills Macondo, But Its ...
blogs.wsj.com/source/2010/09/20/bp-kills-macondo-but-its-legacy-lives-on/
THE WALL STREET JOURNAL.
THE SOURCE
SEPTEMBER 20, 2010, 12:44 PM GMT
BP Kills Macondo, But Its Legacy Lives On
By James Herron
BP confirmed late Sunday that the Macondo well that leaked almost five million barrels of oil into the Gulf of Mexico has been permanently sealed, but the well will continue to affect BP and the wider oil industry for many years.
Fire boat response crews battled the blazing remnants of the off shore oil rig Deepwater Horizon on April 21, 2010.
Allen: Well is dead, but m...
edition.cnn.com/2010/US/09/20/gulf.oil.disaster/
CNN
Allen: Well is dead, but much Gulf Coast work remains
By the CNN Wire Staff
September 20, 2010 -- Updated 1317 GMT (2117 HKT)
Click to play
What next for Gulf oil spill?
STORY HIGHLIGHTS
(CNN) -- The ruptured Macondo well, a mile under the Gulf of Mexico off the Louisiana coast, has been pronounced dead
BP oil spill cost hits nearl...
www.guardian.co.uk/environment/2010/sep/20/bp-oil-spill-dee
guardian.co.uk
Business BP
BP oil spill cost hits nearly $10bn
BP has set up a $20bn compensation fund after the Deepwater Horizon disaster, which has so far paid out 19,000 claims totalling more than $240m
Julia Kollewe
guardian.co.uk, Monday 20 September 2010 08.33 BST
BP's costs for the Deepwater Horizon disaster have hit $10bn. Photograph: Ho/Reuters

Organize computing clusters



Social network analysis



Astronomical data analysis

Of the following examples, which would you address using an <u>unsupervised</u> learning algorithm? (Check all that apply.)

Given email labeled as spam/not spam, learn a spam filter.

Given a set of news articles found on the web, group them into set of articles about the same story.

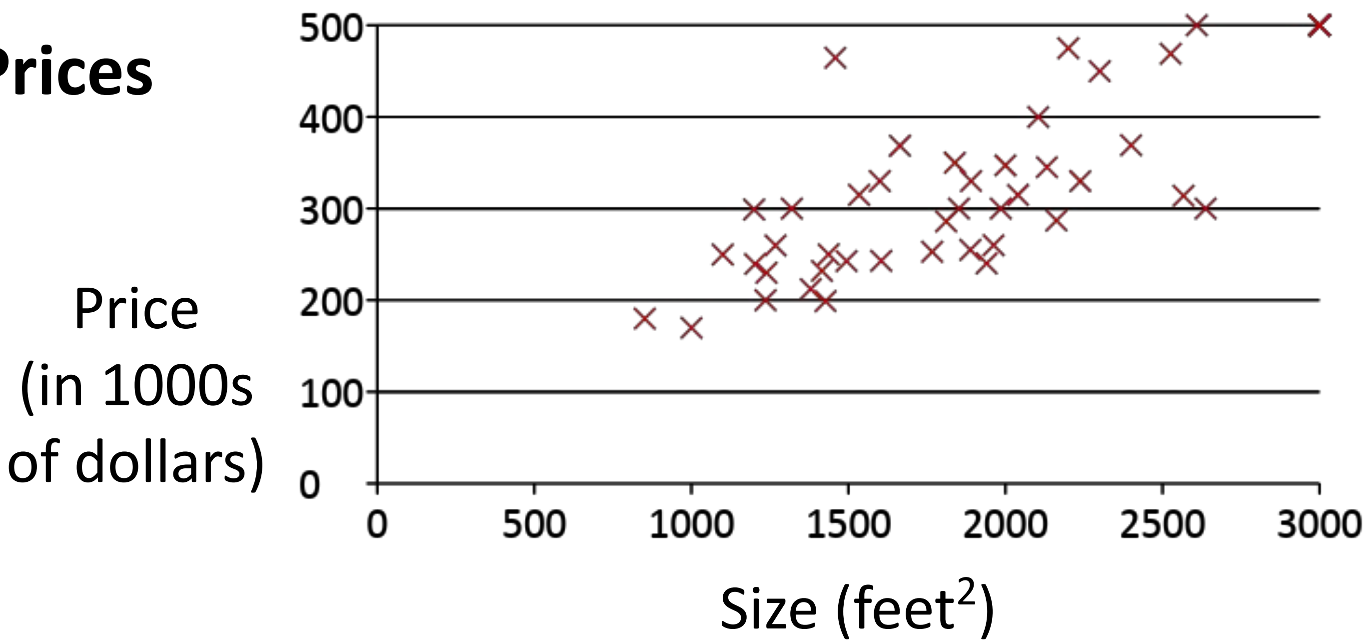
Given a database of customer data, automatically discover market segments and group customers into different market segments.

Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Linear regression with one variable

Model representation

Housing Prices



Supervised Learning

Given the "right answer" for each example in the data.

Regression Problem

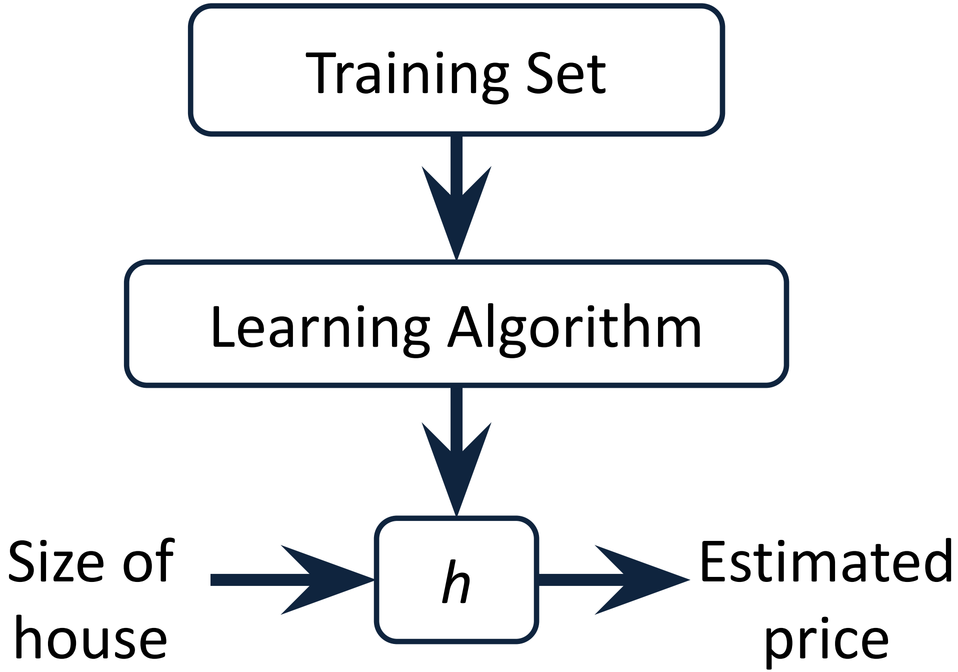
Predict real-valued output

Training set of housing prices

Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

Notation:

m = Number of training examples
x's = "input" variable / features
y's = "output" variable / "target" variable



How do we represent *h* ?

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Linear regression with one variable.
Univariate linear regression.

Linear regression with one variable

Cost function

Training Set

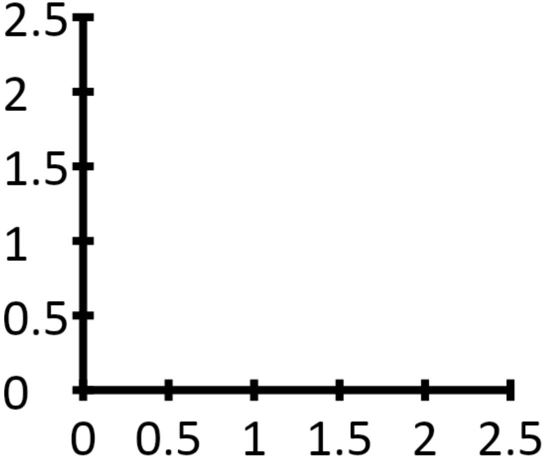
Size in feet^2 (x)	**Price ($) in 1000's (y)**
2104	460
1416	232
1534	315
852	178
…	…

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

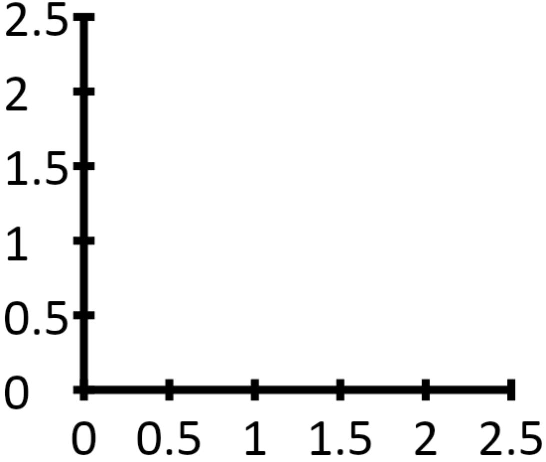
θ_i's: Parameters

How to choose θ_i's ?

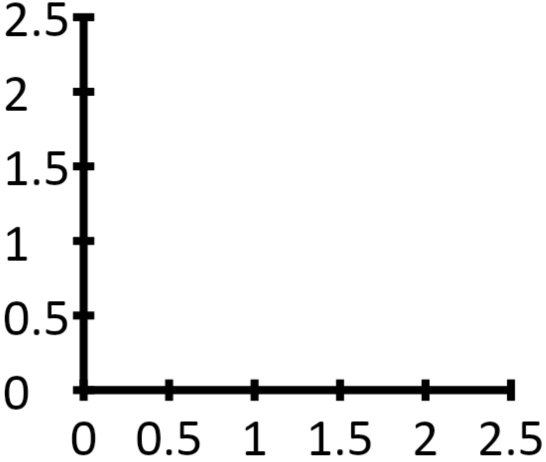
$$h_\theta(x) = \theta_0 + \theta_1 x$$



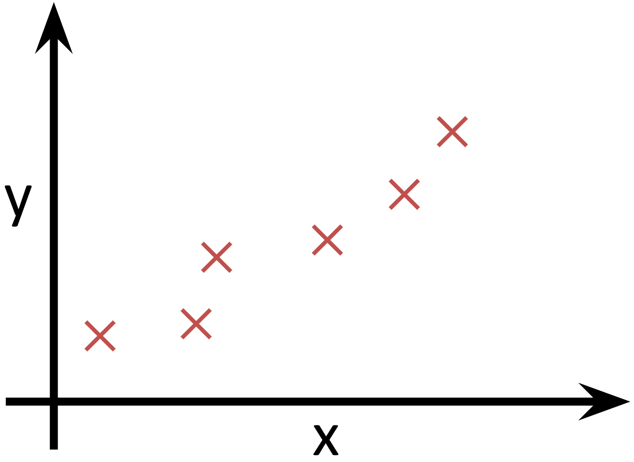
$$\theta_0 = 1.5$$
$$\theta_1 = 0$$



$$\theta_0 = 0$$
$$\theta_1 = 0.5$$



$$\theta_0 = 1$$
$$\theta_1 = 0.5$$



Cost Function
(Squared Errror function):

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}} \; J(\theta_0, \theta_1)$

Idea: Choose θ_0, θ_1 so that $h_\theta(x)$ is close to y for our training examples (x, y)

Linear regression with one variable

Cost function intuition I

Hypothesis:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$

Cost Function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

<u>Simplified (to understand cost function better)</u>

$$h_\theta(x) = \theta_1 x$$

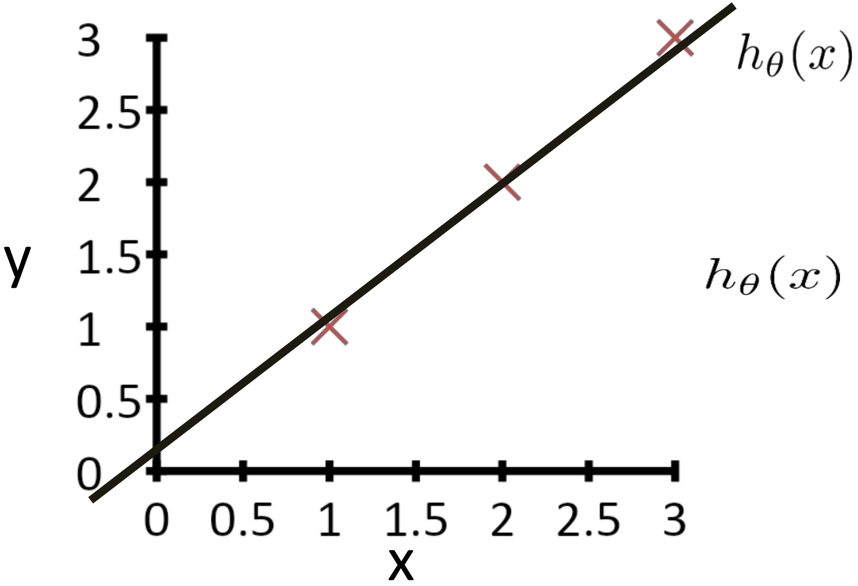
$$\theta_1$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

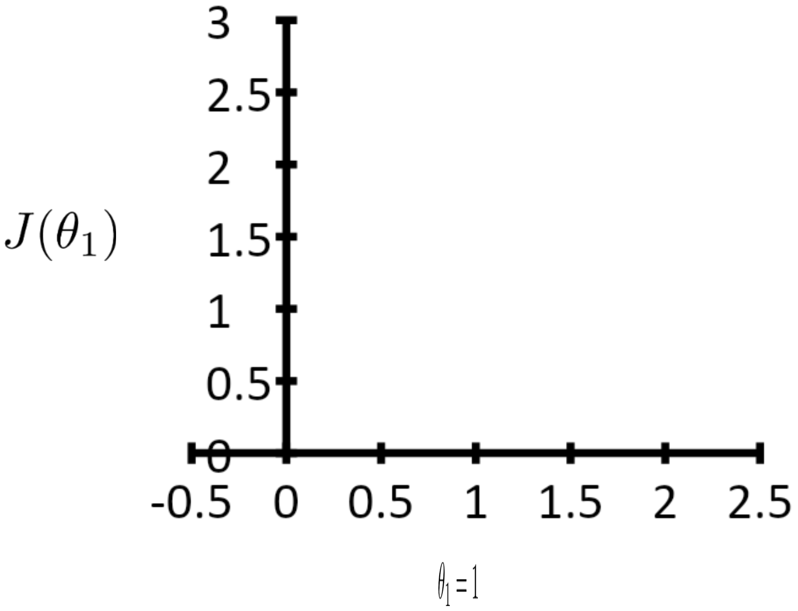
$h_\theta(x)$

(for fixed $\theta_1=1$, this is a function of x)



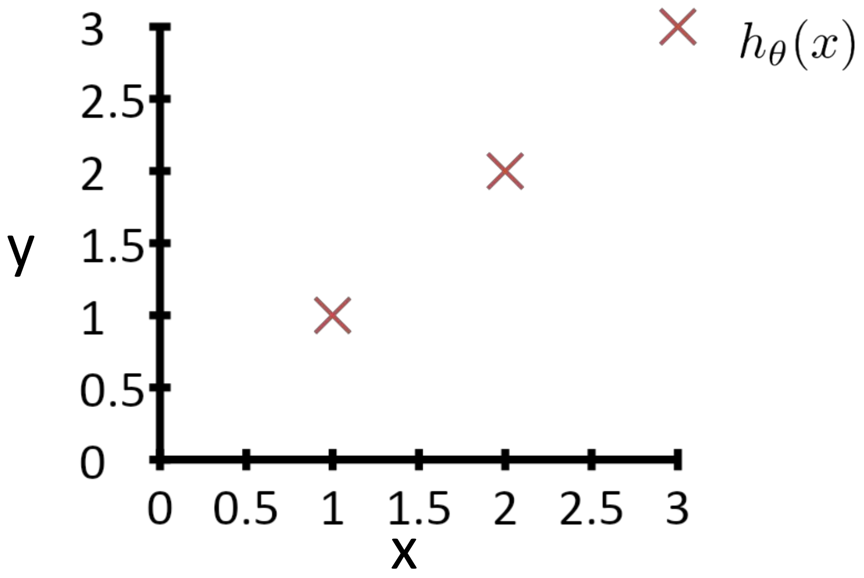
$J(\theta_1)$

(function of the parameter $\theta_1=1$)



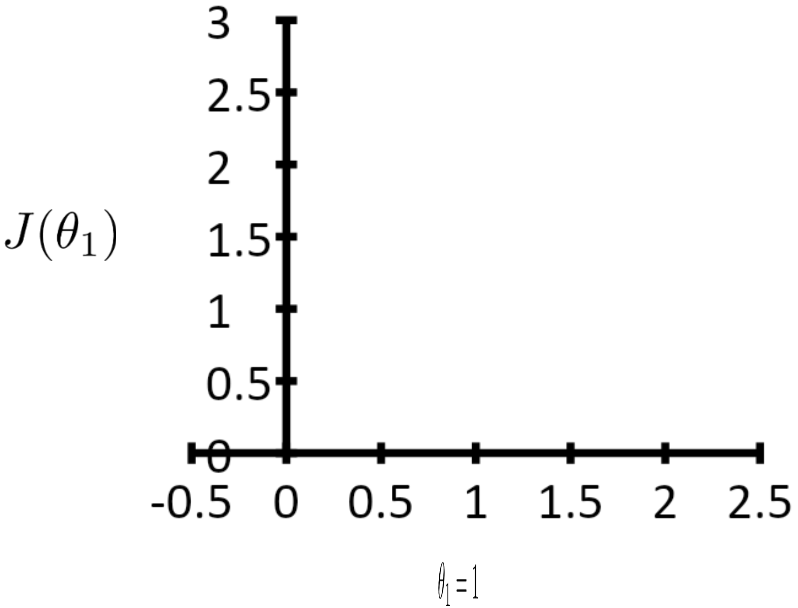
$h_\theta(x)$

(for fixed θ_1=1, this is a function of x)



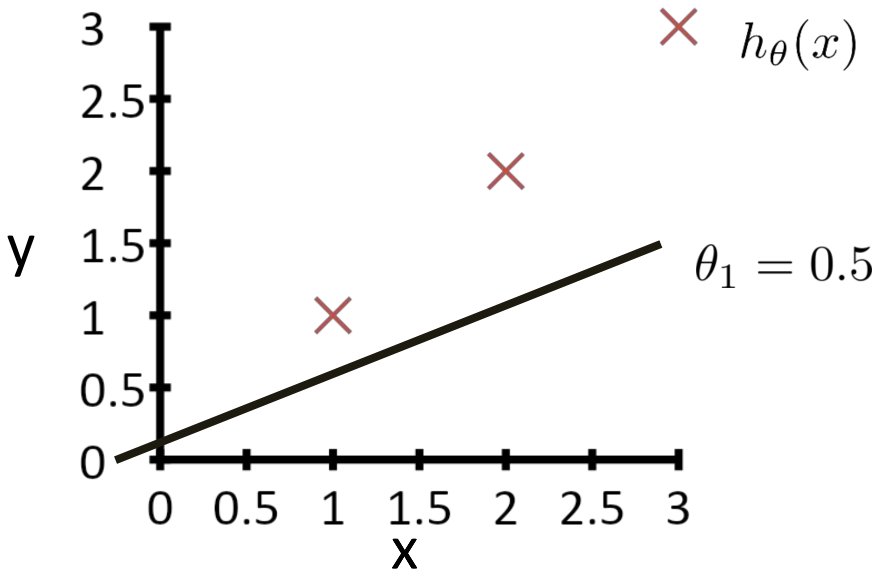
$J(\theta_1)$

(function of the parameter θ_1=1)



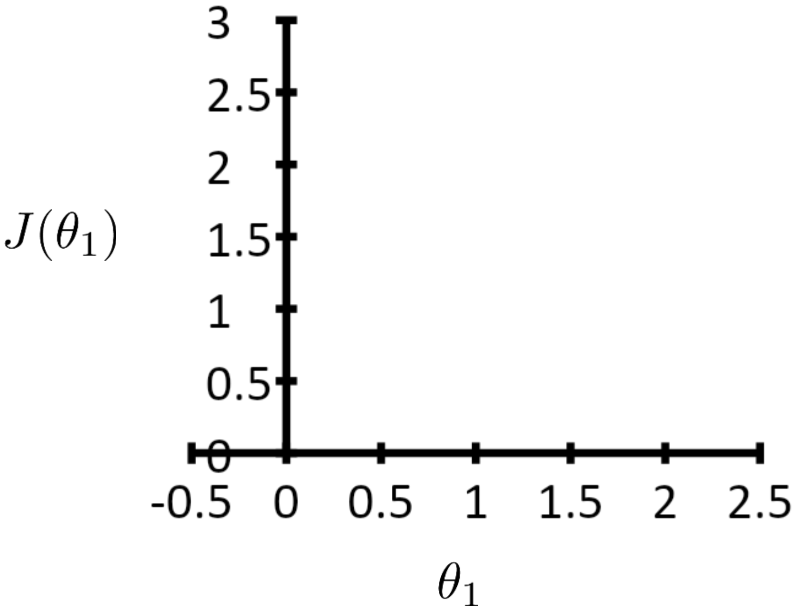
$h_\theta(x)$

(for fixed $\theta_1 = 0.5$, this is a function of x)



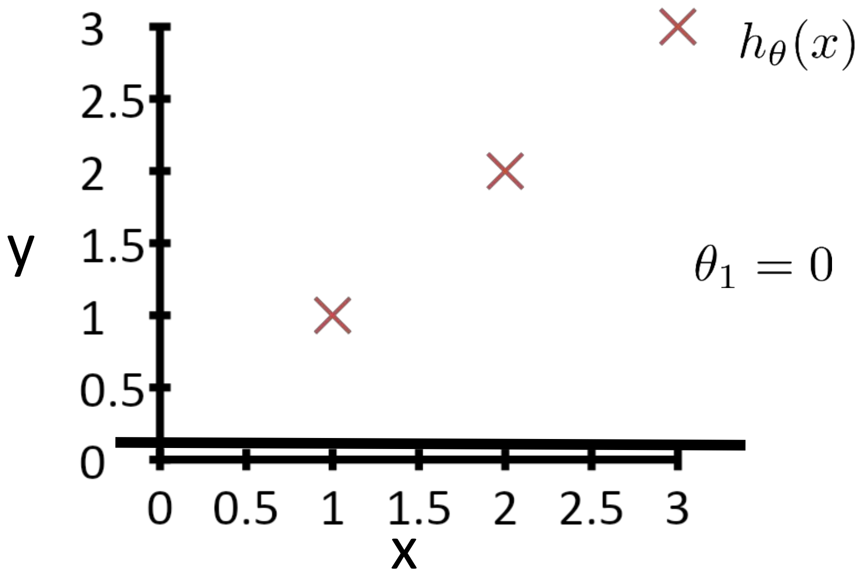
$J(\theta_1)$

(function of the parameter θ_1)



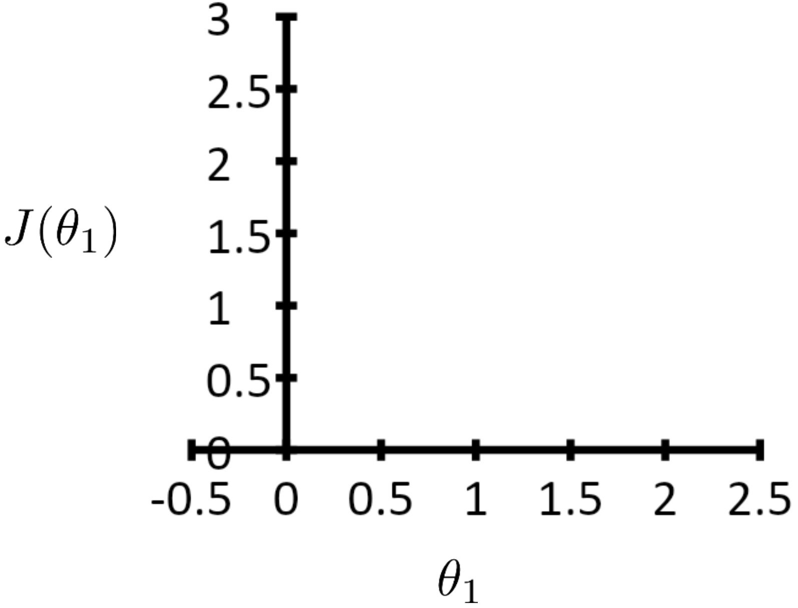
$h_\theta(x)$

(for fixed $\theta_1 = 0$, this is a function of x)



$J(\theta_1)$

(function of the parameter θ_1)



Linear regression with one variable

Cost function intuition II

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

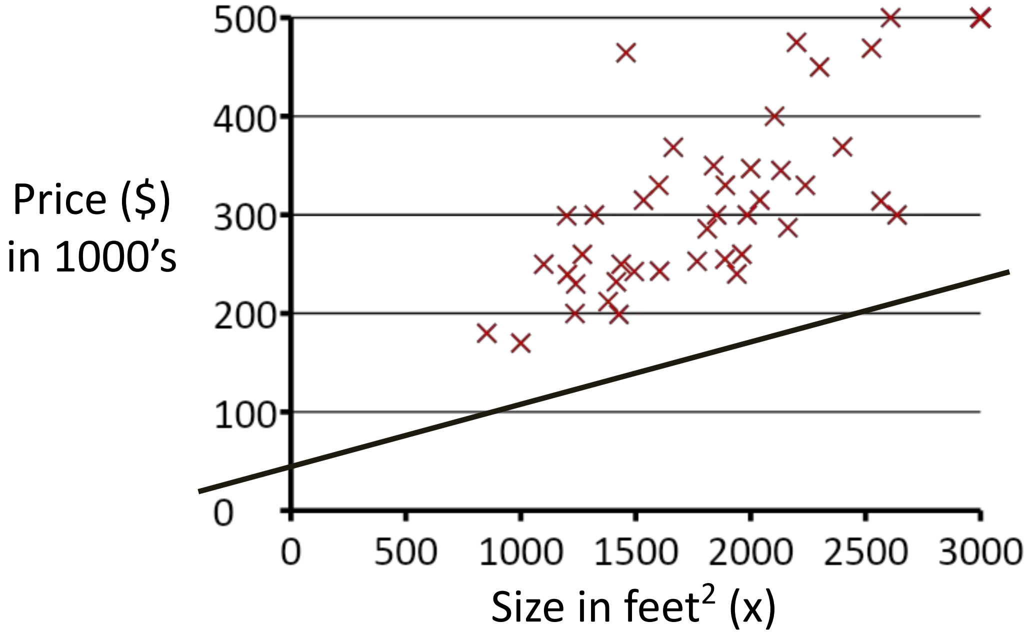
Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

$h_\theta(x)$

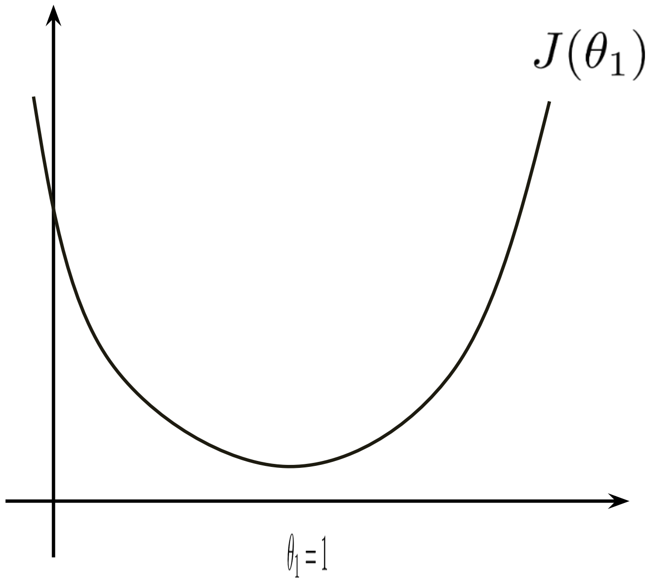
(for fixed θ_0, θ_1, this is a function of x)



θ_0 = 50 θ_1 = 0.06

$J(\theta_0, \theta_1)$

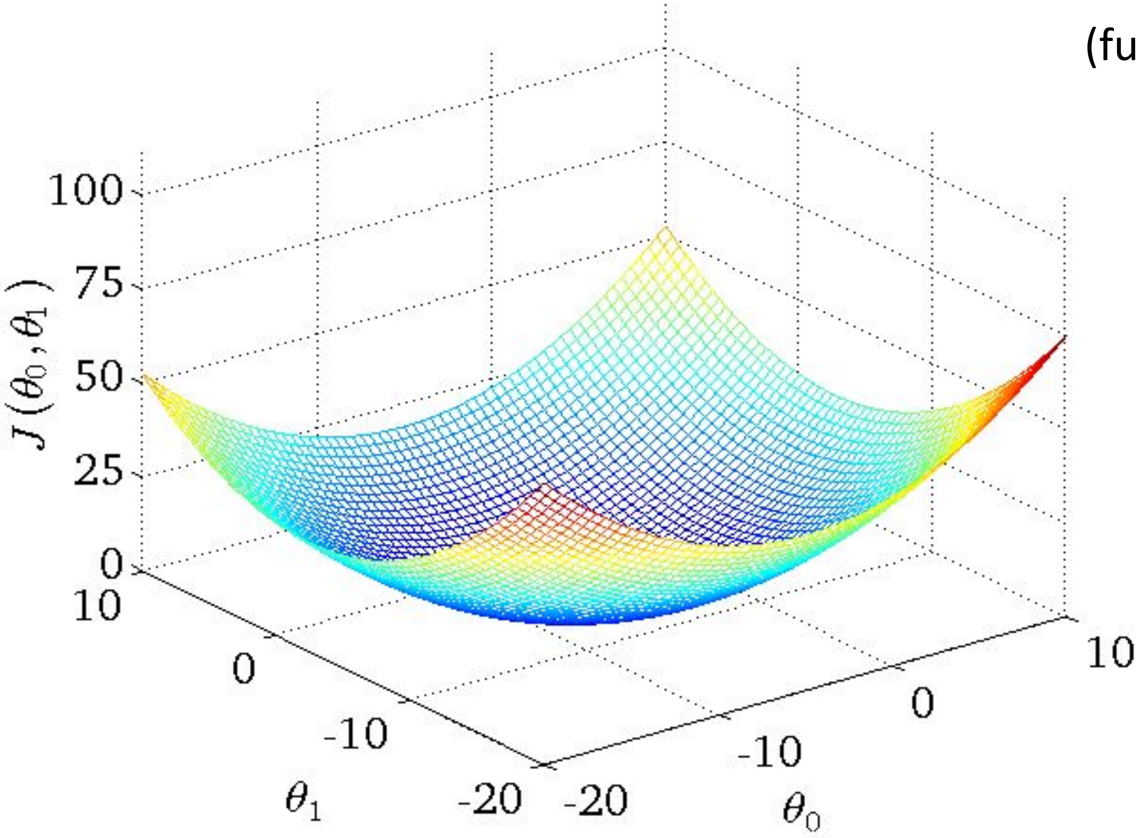
(function of the parameters θ_0, θ_1)



$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

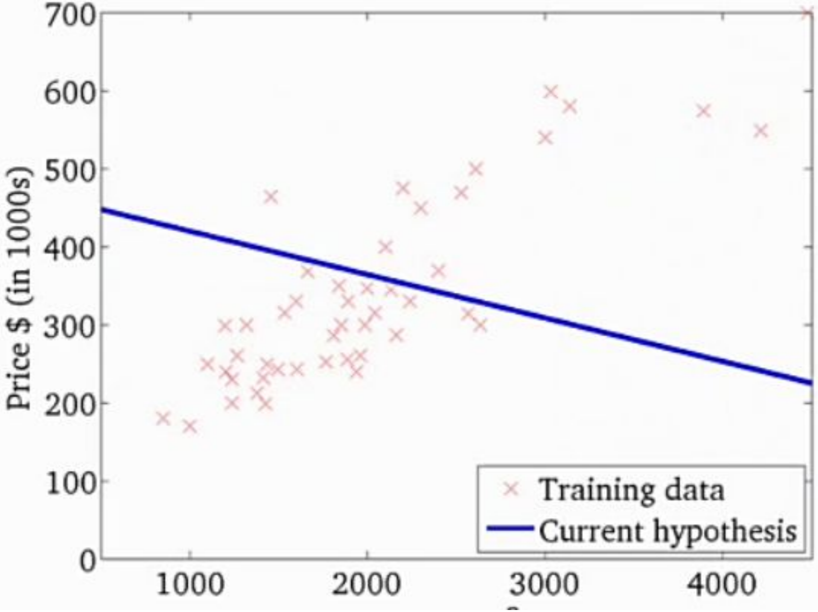
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



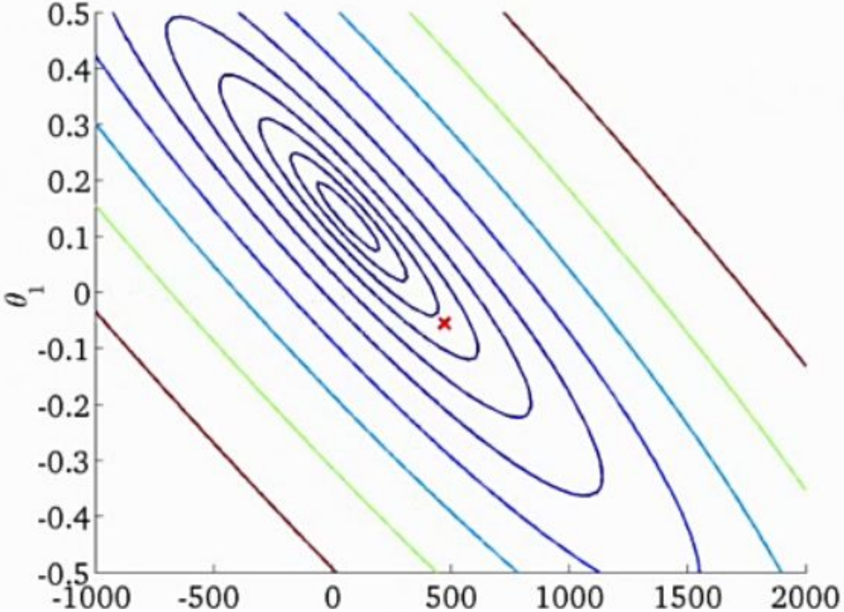
$$h_\theta(x)$$

(for fixed θ_0, θ_1, this is a function of x)



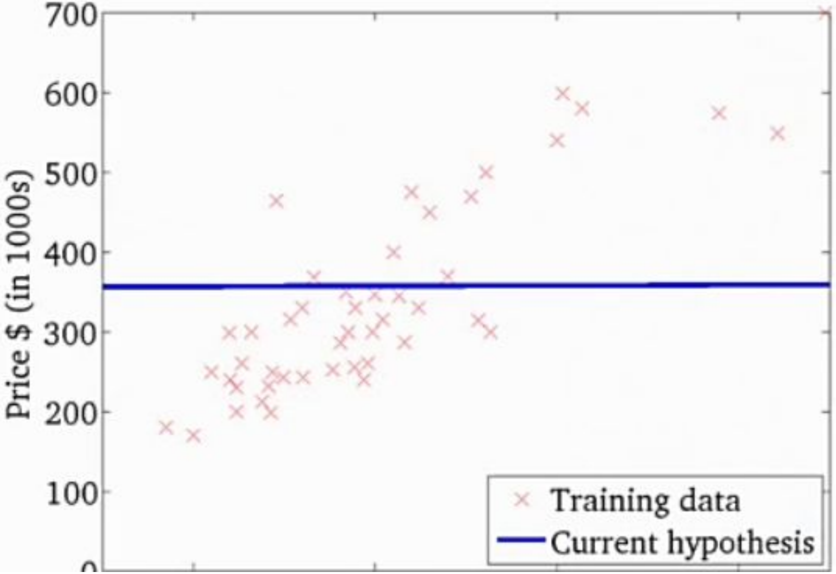
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



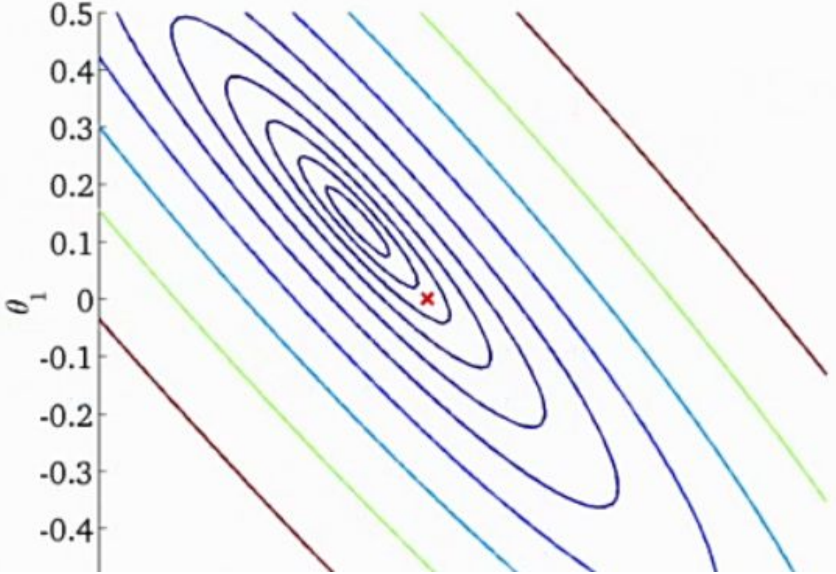
$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



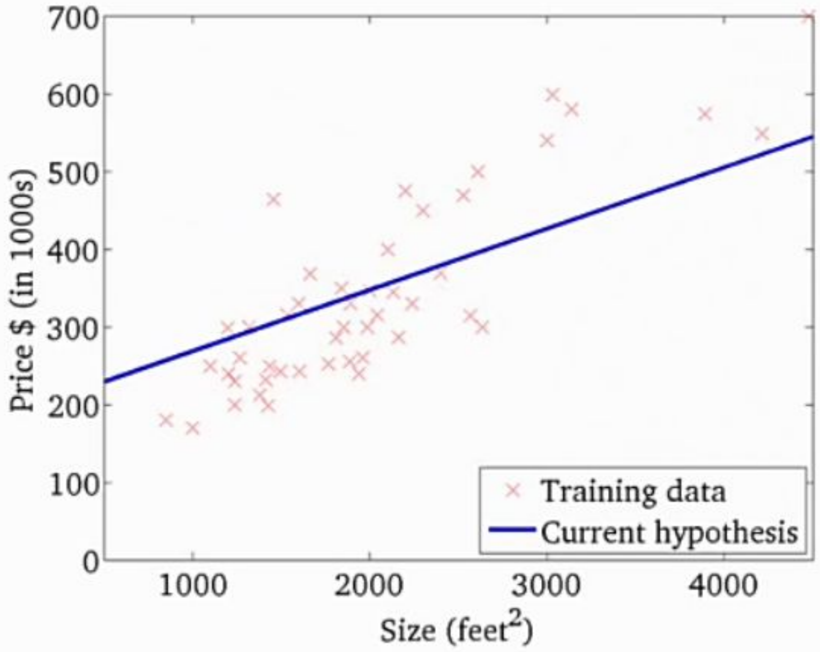
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)



$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)