CPELEC1 0

Machine Intelligence Introduction

Melvin Cabatuan

Electronics & Communications Engineering
De La Salle University–Manila

github.com/melvincabatuan

What is Machine Intelligence [1]?

Home > English Dictionary > machine intelligence

Your search found: machine intelligence / Usage examples / Trends

machine intelligence

Definitions

noun

(British, rare) another term for artificial intelligence

[1] http://www.collinsdictionary.com/dictionary/english/machine-intelligence

What is Artificial Intelligence (AI)?

``It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." - John McCarthy

What is Intelligence?

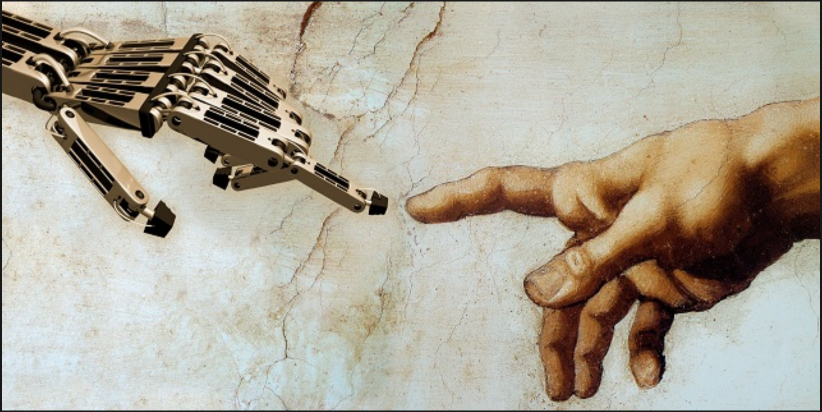
``Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines." - John McCarthy

When did AI research start?

``After WWII, a number of people independently started to work on intelligent machines. The English mathematician Alan Turing may have been the first. He gave a lecture on it in 1947. He also may have been the first to decide that AI was best researched by programming computers rather than by building machines. By the late 1950s, there were many researchers on AI, and most of them were basing their work on programming computers." - John McCarthy

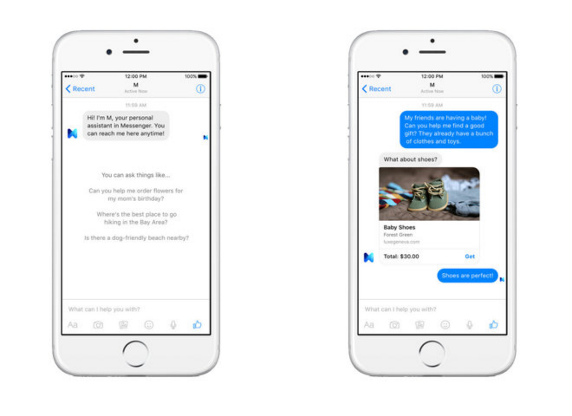
Does AI aim at human-level intelligence?

``Yes. The ultimate effort is to make computer programs that can solve problems and achieve goals in the world as well as humans. However, many people involved in particular research areas are much less ambitious.'' - John McCarthy



Facebook's 'M' virtual assistant

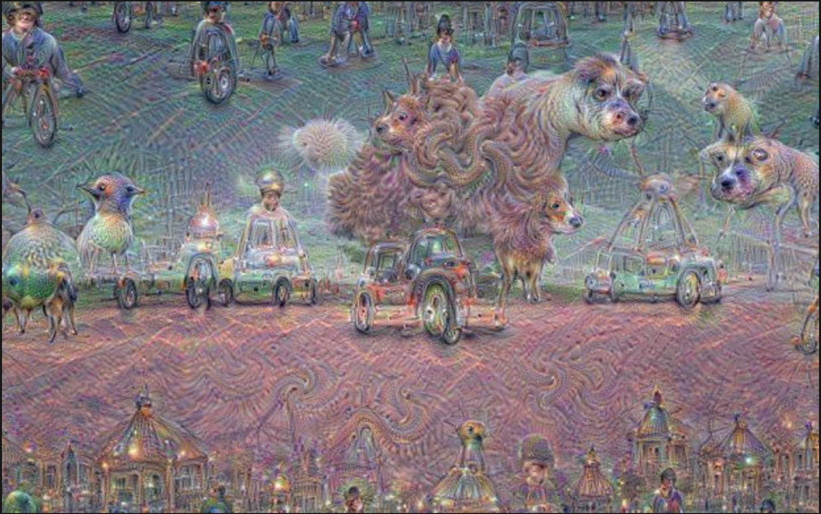
A text-based virtual assistant integrated with Facebook Messenger that unlike Siri, Cortana, or Google Now relies on both human and artificial intelligence.



[1] http://appleinsider.com/articles/15/08/26/facebook-launches-m-virtual-assistant-driven-by-both-human-and-artificial-intelligence

Google's Deepdream

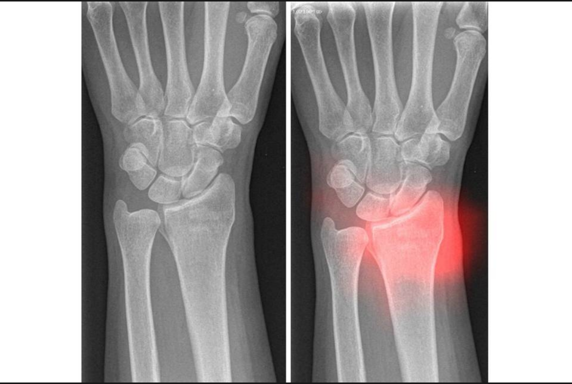
Artificial intelligence meets hallucinations with Google's deep dream.



[1] https://github.com/google/deepdream

IBM's Watson software in Medicine

IBM intends to use around 30 billion images, including X-rays, computerized tomography, and magnetic-resonance-imaging scans to “train” its Watson software to identify ailments such as cancer and heart disease.



[1] http://www.wsj.com/articles/ibm-crafts-a-role-for-artificial-intelligence-in-medicine

AI Sub-fields

- Bioinformatics/ Genomics
- Computer Vision
- Machine Learning
- Natural language processing
- Reasoning/Logic
- Robotics
- Sensor and Network

Computer Vision

It is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. - Wikipedia

Robotics

It refers to the design, construction, operation, and application of robots. The term was coined by the Russian-born American scientist and writer Isaac Asimov.

Laws of robotics

- Law Zero: ``A robot may not injure humanity, or, through inaction, allow humanity to come to harm."
- Law One: ``A robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate a higher order law."
- Law Two: ``A robot must obey orders given it by human beings, except where such orders would conflict with a higher order law."
- Law Three: ``A robot must protect its own existence as long as such protection does not conflict with a higher order law."

Machine Learning

``Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel.

Machine Learning

``A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E'' - Tom Mitchell.

Question

"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?

- ◯ Classifying emails as spam or not spam.
- ◯ Watching you label emails as spam or not spam.
- ◯ The number (or fraction) of emails correctly classified as spam/not spam.

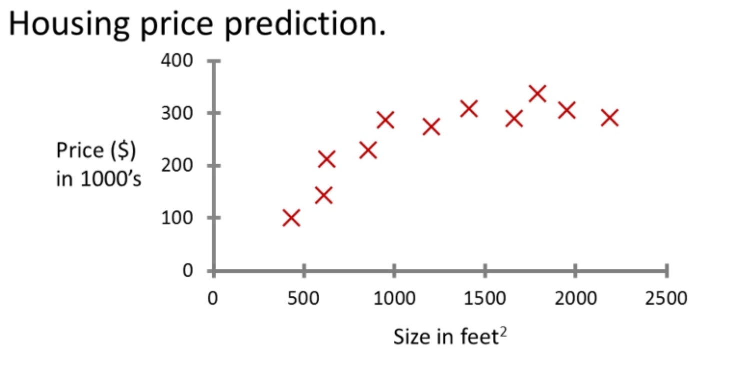
[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Discussion questions

- Will machine intelligence surpass human intelligence?
- What are the most important components of machine intelligence?
- Differentiate Supervised learning from Unsupervised learning.

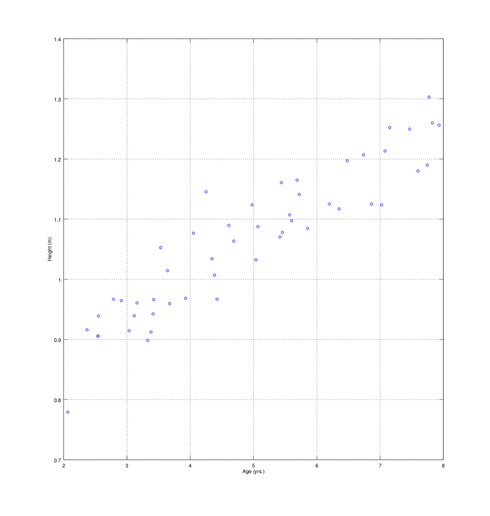
Supervised Learning

``Supervised learning entails learning a mapping between a set of input variables X and an output variable Y and applying this mapping to predict the outputs for unseen data."



[1]Cunningham, P. et al. "Supervised learning." Machine Learning Techniques for Multimedia. Springer Berlin Heidelberg, 2008. 21-49.
[2]Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Ex. Heights for various boys between the ages 2-8 yrs.

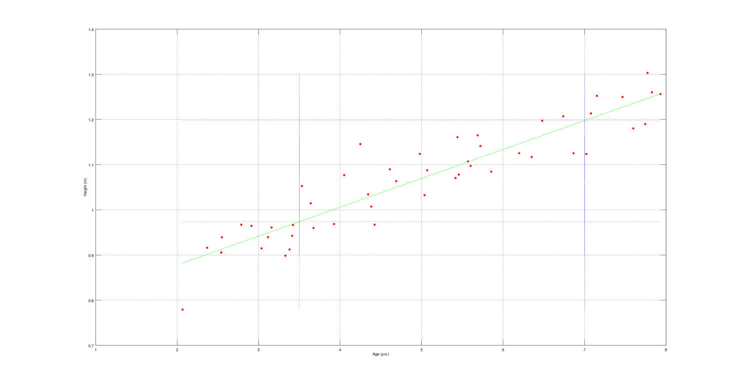


[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Linear Regression Model

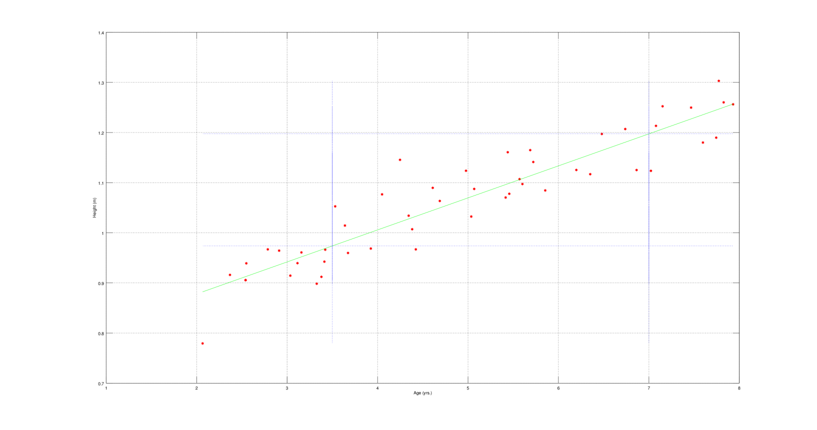
$$h_\theta(x) = \theta^T x = \sum_{i=0}^{n} \theta_i x_i$$

Linear regression fits a line to a set of data.



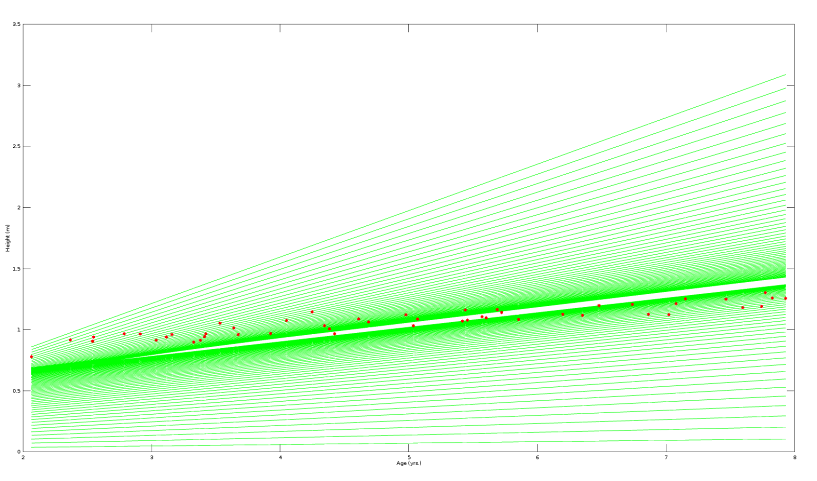
Linear Regression Model

Height prediction of boys at ages 3.5 and 7 are 0.9737 and 1.197 m.



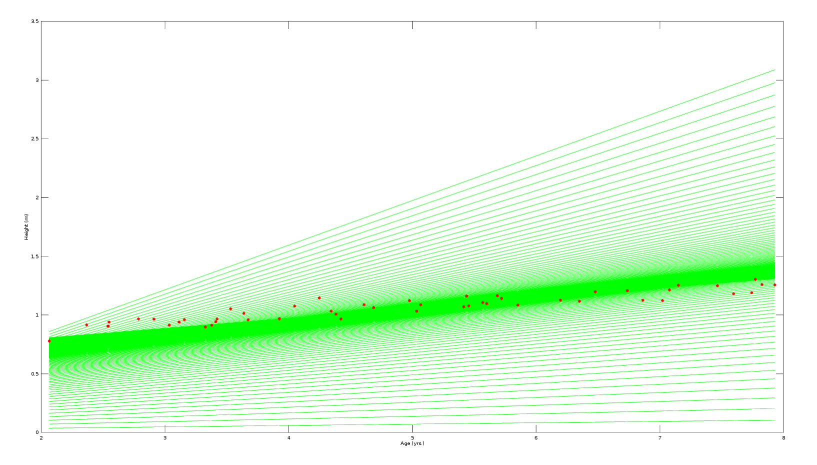
Gradient Descent

Line predictions for every iteration



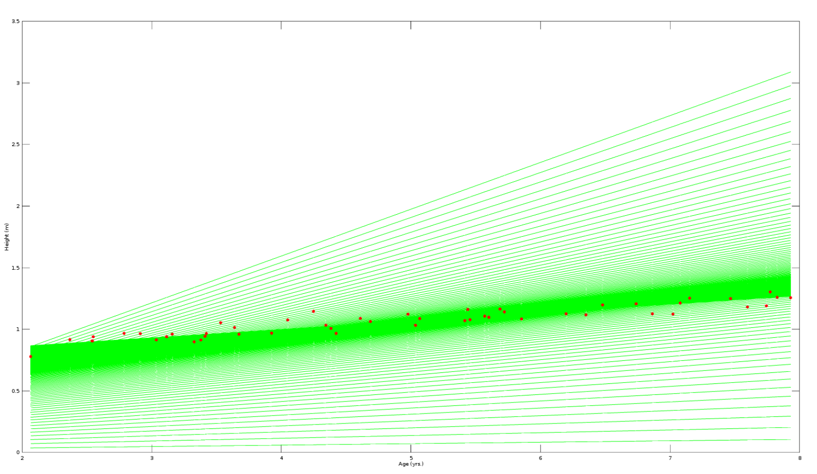
Gradient Descent

Line predictions for every iteration



Gradient Descent

Line predictions for every iteration



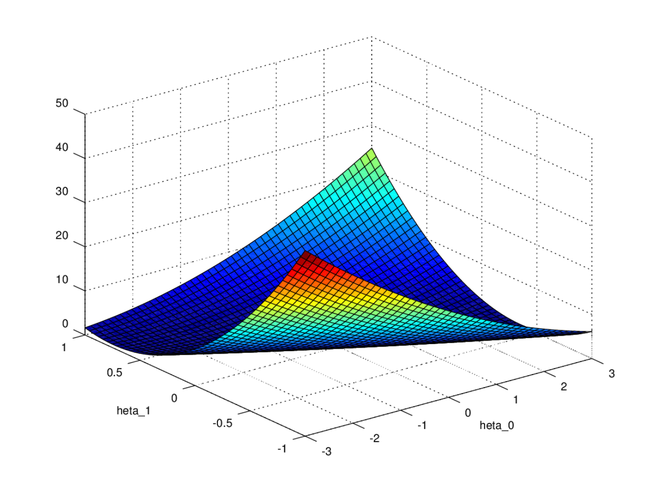
How to find the best line hypothesis, h_θ?

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

- Define a cost function, $J(\theta)$, that measures how 'bad' a given line is.
- Iterate through each data point, $(x^{(i)}, y^{(i)})$, and measure the distance(error) between the current hypothesis, h_θ, to the data points.
- It's conventional to utilize the squared error to ensure that it is positive and to make the error function differentiable.

Cost Function Surface Plot

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



Octave source code

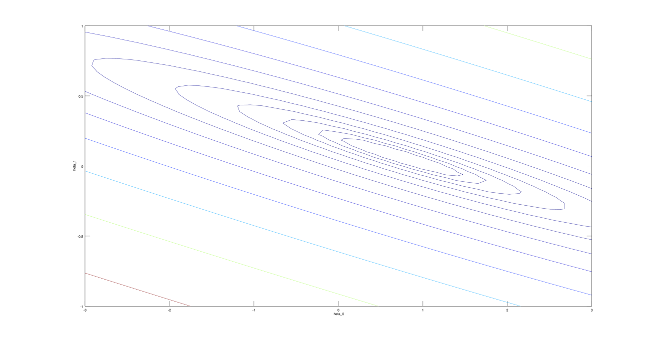
Gradient Descent batch update

```
for i = 1:length(theta0_vals)
  for j = 1:length(theta1_vals)
     t = [theta0_vals(i); theta1_vals(j)];
      J_vals(i,j) = (1/(2*m)) .* (x * t - y)' *
         (x * t - y);
    endfor
endfor

J_vals = J_vals';

figure, contour(theta0_vals, theta1_vals,
   J_vals, logspace(-1.5,1.5,12))
```

[1] github.com/melvincabatuan/mlintroduction/blob/master/CostFunctionCor

Cost Function Contour Plot

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



Understanding the Cost function plot

- Each point in the previous two-dimensional space represents a hypothesis, h_θ, which is a line in this case.
- The height of the function, $J(\theta)$, at each point is the error value for that particular hypothesis.
- Thus, the main goal for our algorithm in minimizing the Cost function, $J(\theta)$, is to move downhill in each step and eventually reach the lowest point.

Minimizing the Cost function

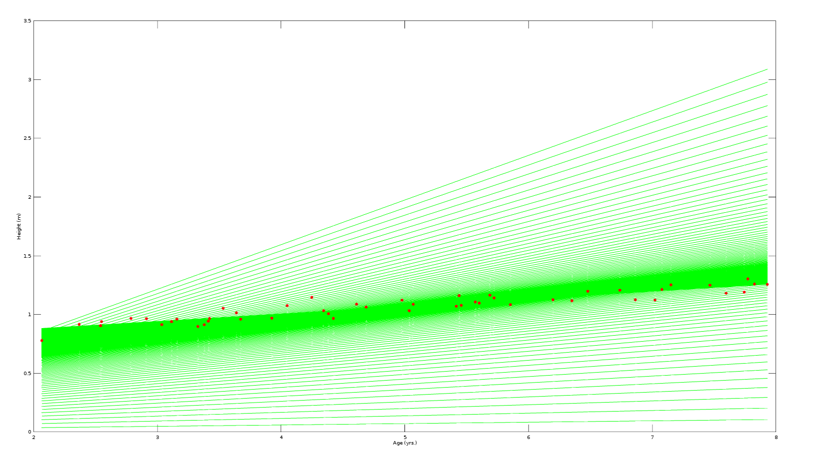
$$\underset{\theta_0,\theta_1}{\text{minimize}} \quad J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

- To compute the minimum, we need to differentiate the cost function.
- In particular, we have to compute a partial derivative with respect to each model parameter, θ_0, θ_1, and so on.
- Gradient:

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

Gradient Descent

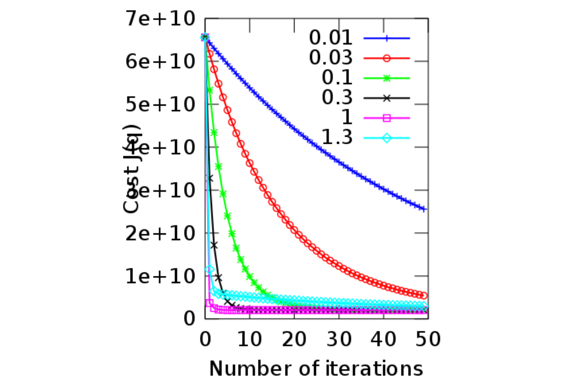
$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$



Understanding the learning parameter, α

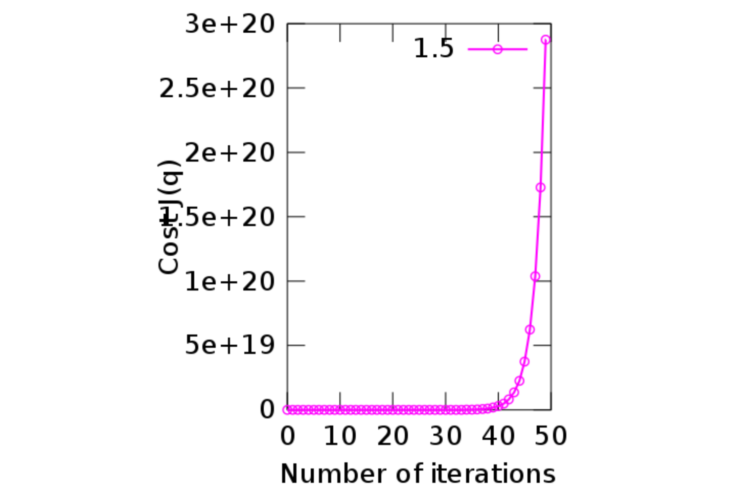
α controls how large of a step we take downhill during each iteration.

If we take too small steps, it will require many iterations to arrive at the minimum resulting to a slow convergence.



Understanding the learning parameter, α

If we take too large of a step, we may step over the minimum, thus, it may not converge to the minimum.



Octave source code

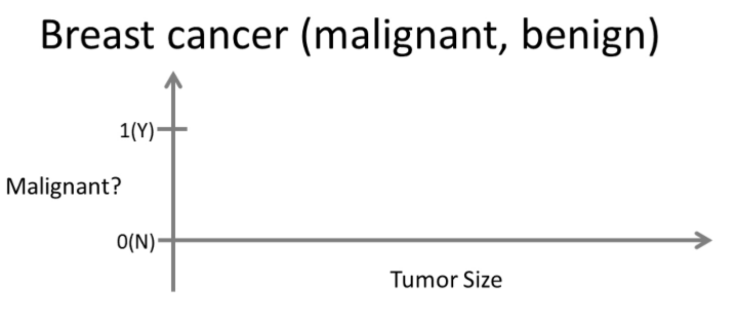
Gradient Descent batch update

```
 for ii = 1:MAX_ITERATION
       % Dimensions: 2x50 (50x2 * 2x1 - 50x1)
          = 2x1
       gradient = (1/m).* x' * ((x * theta) -
          y);
       theta = theta - alpha .* gradient;
endfor
```

[1] github.com/melvincabatuan/mlintroduction/blob/master/GradientDescent

Supervised Learning

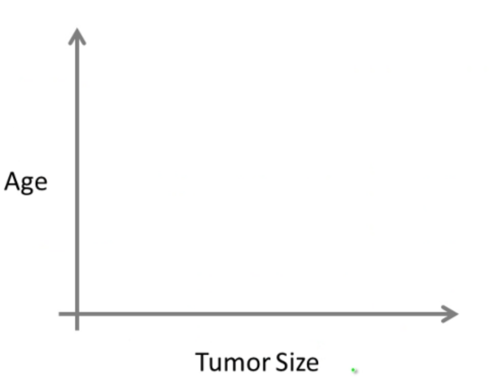
``Supervised learning entails learning a mapping between a set of input variables X and an output variable Y and applying this mapping to predict the outputs for unseen data."



[1] Cunningham, P. et al. "Supervised learning." Machine Learning Techniques for Multimedia. Springer Berlin Heidelberg, 2008. 21-49.
[2] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Supervised Learning

``Supervised learning entails learning a mapping between a set of input variables X and an output variable Y and applying this mapping to predict the outputs for unseen data."



[1]Cunningham, P. et al. "Supervised learning." Machine Learning Techniques for Multimedia. Springer Berlin Heidelberg, 2008. 21-49.
[2]Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Question

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?

- ○ Treat both as classification problems.
- ○ Treat problem 1 as a classification problem, problem 2 as a regression problem.
- ○ Treat problem 1 as a regression problem, problem 2 as a classification problem.
- ○ Treat both as regression problems.

[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Binary Classification Problem

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

- 0: "Negative Class" (E.x. benign tumor, not spam, not fraudulent)
- 1: ``Positive Class" (E.x. malignant tumor, spam, fraudulent)
- Thus, $0 \leq h_\theta(x) \leq 1$.

Hypothesis Representation

$$h_\theta(x) = g(\theta^T x)$$

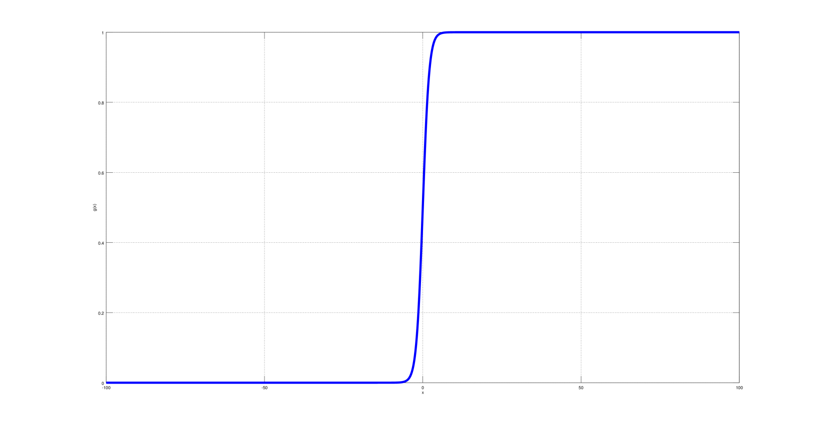
$$g(z) = \frac{1}{1 + e^{-z}}$$

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

- Instead of a linear regression line, the classification hypothesis takes the form of a Sigmoid or Logistic function.
- Sigmoid function satisfies the constraint:

$$0 \leq h_\theta(x) \leq 1$$

Sigmoid or Logistic Function



Octave source code

Plotting Sigmoid function

```
g = inline('1.0 ./ (1.0 + exp(-z))');
% Usage: g(x), e.g. g(5)
x = linspace(-100, 100, 1000);
plot(x,g(x),'linewidth', 5);
xlabel('x'), ylabel('g(x)');
```

[1] github.com/melvincabatuan/mlintroduction

Hypothesis Output Interpretation

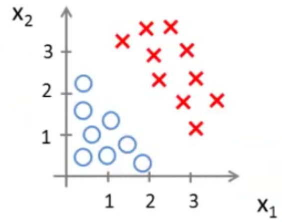
$$h_\theta(x) = P(y = 1|x; \theta)$$

- $h_\theta(x) =$ estimated probability that $y = 1$ on input x
- E.x. If $h_\theta(x) = 0.7$, tell patient that 70 % chance of tumor being malignant
- Thus, if $h_\theta(x) \geq 0.5$, then predict "$y = 1$" and if $h_\theta(x) < 0.5$ then predict ``$y = 0$``
- What values of $\theta^T x$ does the hypothesis predict "$y = 1$"? How about ``$y = 0$"?

Seatwork 1

Linear Decision Boundary

Given the hypothesis function, $h_\theta(x)$, and decision boundary line, determine the model parameters. θ.
When will the hypothesis predict ``$y = 1$" ? ``$y = 0$"?

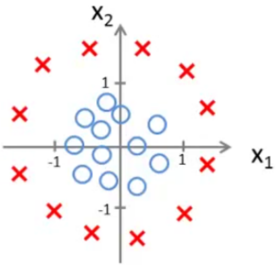


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Seatwork 2

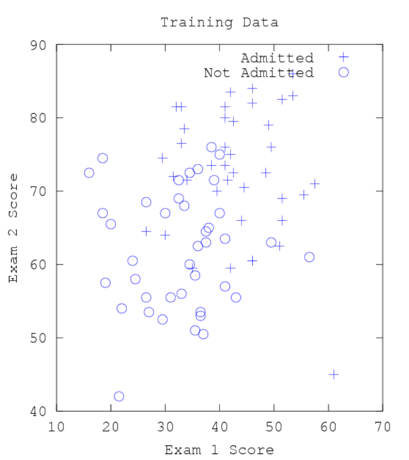
Nonlinear Decision Boundary

Given the hypothesis function,$h_\theta(x)$, and decision boundary line, determine the model parameters. θ. When will the hypothesis predict ``$y = 1$'' ? ``$y = 0$''?



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

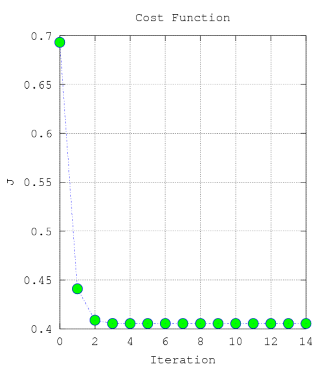
Logistic Regression Example Data



[1] http://melvincabatuan.github.io/Machine-Learning-Activity-3
[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

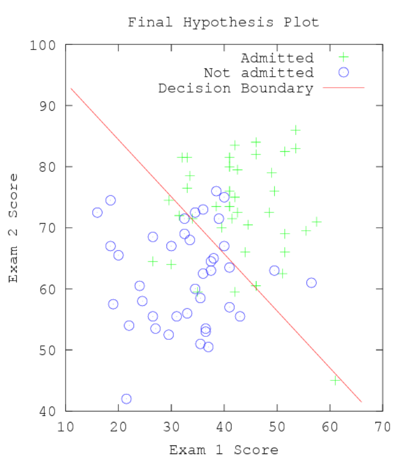
Logistic Regression Cost Function

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



[1] http://melvincabatuan.github.io/Machine-Learning-Activity-3
[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Decision Boundary



[1] http://melvincabatuan.github.io/Machine-Learning-Activity-3
[1] Andrew Ng Lecture, https://class.coursera.org/ml-005/lecture

Minimizing the Cost function

$$\underset{\theta}{\text{minimize}} \quad J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[-y^{(i)} \log(h_\theta(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right]$$

- To compute the minimum, we need to differentiate the cost function.
- In particular, we have to compute a derivative with respect to the model parameter, θ.
- Gradient:

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

Appendix

Inserting source code

```
#include<stdio.h>
#include<iostream>
// A comment
int main(void)
{
printf("Hello World\n");
return 0;
}
```