Introduction

Welcome
Sachin Joshi –
Research Computing

Machine Learning
Contents

• Introduction
• Machine Learning Definition
• Supervised ML
• Unsupervised ML
• Linear Regression with one variable
  • Model Representation
  • Cost Function
  • Next Seminar – Gradient Descent Algorithm
iPhoto ‘11

From your Facebook Wall to your coffee table to your best friend’s inbox (or mailbox). Do more with your photos than you ever thought possible. And do it all in one place. iPhoto.

Watch the iPhoto video ▶

What’s New in iPhoto

What is iPhoto?
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.
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.
- Self-customizing programs
  E.g., Amazon, Netflix product recommendations
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.
- Self-customizing programs
  E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).
Introduction

What is machine learning

Machine Learning
Machine Learning definition
Machine Learning definition

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

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

- 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.
- None of the above—this is not a machine learning problem.
“A computer program is said to \textit{learn} from experience \(E\) with respect to some task \(T\) and some \textit{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.
- None of the above—this is not a machine 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.”

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 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.
- None of the above—this is not a machine learning problem.
Machine learning algorithms:
- Supervised learning
- Unsupervised learning

Others: Reinforcement learning, recommender systems.

Also talk about: Practical advice for applying learning algorithms.
Introduction
Supervised Learning
Machine Learning
Housing price prediction.

Supervised Learning: “right answers” given

Regression: Predict continuous valued output (price)
Breast cancer (malignant, benign)

Classification
Discrete valued output (0 or 1)

Malignant? 1(Y) 0(N)

Tumor Size

Tumor Size

benign type 1 cancer
- 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?

- 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.
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.
Introduction
Unsupervised Learning
Machine Learning
Supervised Learning
Unsupervised Learning
BP Oil Well, Site of National Catastrophe, Dies at One

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

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

[Source: Daphne Koller]
Organize computing clusters

Social network analysis

Market segmentation

Astronomical data analysis

Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, Madison)
Cocktail party problem

Speaker #1

Speaker #2

Microphone #1

Microphone #2
Cocktail party problem algorithm

\[ [W, s, v] = \text{svd}(\text{repmat}(\text{sum}(x.*x,1), \text{size}(x,1), 1).*x)*x'); \]

[Source: Sam Roweis, Yair Weiss & Eero Simoncelli]
Of the following examples, which would you address using an unsupervised 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.
Of the following examples, which would you address using an unsupervised 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.
Quiz 1 – ML Basics

1. 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 we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for P?
   o The process of the algorithm examining a large amount of historical weather data.
   o The weather prediction task.
   o The probability of it correctly predicting a future date's weather.
   o None of these.

2. Suppose you are working on weather prediction, and use a learning algorithm to predict tomorrow’s temperature (in degrees Centigrade/Fahrenheit). Would you treat this as a classification or a regression problem?
   o Regression
   o Classification

3. Suppose you are working on stock market prediction, and you would like to predict the price of a particular stock tomorrow (measured in dollars). You want to use a learning algorithm for this. Would you treat this as a classification or a regression problem?
   o Regression
   o Classification

4. Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. Which of the following would you apply supervised learning to? (Select all that apply.) In each case, assume some appropriate dataset is available for your algorithm to learn from.
   qx Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number of groups of essays that are somehow “similar” or “related”.
   qx Examine a large collection of emails that are known to be spam email, to discover if there are sub-types of spam mail.
   qx Given 50 articles written by male authors, and 50 articles written by female authors, learn to predict the gender of a new manuscript’s author (when the identity of this author is unknown).
   qx Given historical data of children’s ages and heights, predict children’s height as a function of their age.

5. Which of these is a reasonable definition of machine learning?
   o Machine learning is the field of allowing robots to act intelligently.
   o Machine learning is the science of programming computers.
   o Machine learning learns from labeled data.
   o Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.
Quiz 1 – ML Basics

1. 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. Supposed we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for P?
   - The process of the algorithm examining a large amount of historical weather data.
   - The weather prediction task.
   - The probability of it correctly predicting a future date's weather.
   - None of these.

2. Suppose you are working on weather prediction, and use a learning algorithm to predict tomorrow's temperature (in degrees Centigrade/Fahrenheit). Would you treat this as a classification or a regression problem?
   - Regression
   - Classification

3. Suppose you are working on stock market prediction, and you would like to predict the price of a particular stock tomorrow (measured in dollars). You want to use a learning algorithm for this. Would you treat this as a classification or a regression problem?
   - Regression
   - Classification

4. Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. Which of the following would you apply supervised learning to? (Select all that apply.) In each case, assume some appropriate dataset is available for your algorithm to learn from.
   - Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number of groups of essays that are somehow "similar" or "related".
   - Examine a large collection of emails that are known to be spam email, to discover if there are sub-types of spam mail.
   - Given 50 articles written by male authors, and 50 articles written by female authors, learn to predict the gender of a new manuscript’s author (when the identity of this author is unknown).
   - Given historical data of children’s ages and heights, predict children’s height as a function of their age.

5. Which of these is a reasonable definition of machine learning?
   - Machine learning is the field of allowing robots to act intelligently.
   - Machine learning is the science of programming computers.
   - Machine learning learns from labeled data.
   - Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.
Linear regression with one variable

Model representation

Machine Learning
Housing Prices (Portland, OR)

Price (in 1000s of dollars)

Supervised Learning
Given the “right answer” for each example in the data.

Regression Problem
Predict real-valued output

Classification: Discrete-valued output
<table>
<thead>
<tr>
<th>Size in feet$^2$ ($x$)</th>
<th>Price ($) in 1000's ($y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2104</td>
<td>460</td>
</tr>
<tr>
<td>1416</td>
<td>232</td>
</tr>
<tr>
<td>1534</td>
<td>315</td>
</tr>
<tr>
<td>852</td>
<td>178</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Notation:

$m = \text{Number of training examples}$

$x$'s = “input” variable / features

$y$’s = “output” variable / “target” variable

$(x, y)$ - one training example

$(x^{(i)}, y^{(i)})$ - $i^{th}$ training example

$m = 47$

\[
\begin{align*}
    x^{(1)} &= 2104 \\
    x^{(2)} &= 1416 \\
    y^{(i)} &= 460
\end{align*}
\]
Consider the training set shown below. \((x^{(i)}, y^{(i)})\) is the \(i^{th}\) training example. What is \(y^{(3)}\)?

<table>
<thead>
<tr>
<th>Size in feet²((x))</th>
<th>Price($) in 1000’s((y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2104</td>
<td>460</td>
</tr>
<tr>
<td>1416</td>
<td>232</td>
</tr>
<tr>
<td>1534</td>
<td>315</td>
</tr>
<tr>
<td>852</td>
<td>178</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>

- 1416
- 1534
- 315
- 0
Consider the training set shown below. \((x^{(i)}, y^{(i)})\) is the \(i^{th}\) training example. What is \(y^{(3)}\)?

<table>
<thead>
<tr>
<th>Size in feet(^2) ((x))</th>
<th>Price($) in 1000’s ((y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2104</td>
<td>460</td>
</tr>
<tr>
<td>1416</td>
<td>232</td>
</tr>
<tr>
<td>1534</td>
<td>315</td>
</tr>
<tr>
<td>852</td>
<td>178</td>
</tr>
<tr>
<td>….</td>
<td>….</td>
</tr>
</tbody>
</table>

- 1416
- 1534
- 315
- 0
How do we represent $h$?

Linear regression with one variable.
Univariate linear regression.

$h(x) = \theta_0 + \theta_1 x$

Shorthand: $h(x)$
Linear regression with one variable

Cost function

Machine Learning
Training Set

<table>
<thead>
<tr>
<th>Size in feet(^2) (x)</th>
<th>Price ($) in 1000's (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2104</td>
<td>460</td>
</tr>
<tr>
<td>1416</td>
<td>232</td>
</tr>
<tr>
<td>1534</td>
<td>315</td>
</tr>
<tr>
<td>852</td>
<td>178</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\[ m = 47 \]

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

\( \theta_i \)'s: Parameters

How to choose \( \theta_i \)'s?
\[ h_\theta(x) = \theta_0 + \theta_1 x \]

- \( h(x) = 1.5 + 0 \cdot x \)
- \( \theta_0 = 1.5 \)
- \( \theta_1 = 0 \)

- \( h(x) = 0.5x \)
- \( \theta_0 = 0 \)
- \( \theta_1 = 0.5 \)

- \( h_\theta(x) \)
- \( \theta_0 = 1 \)
- \( \theta_1 = 0.5 \)
Consider the plot below of $h_\theta(x) = \Theta_0 + \Theta_1 x$. What are $\Theta_0$ and $\Theta_1$?

- $\Theta_0 = 0$, $\Theta_1 = 1$
- $\Theta_0 = 0.5$, $\Theta_1 = 1$
- $\Theta_0 = 1$, $\Theta_1 = 0.5$
- $\Theta_0 = 1$, $\Theta_1 = 1$
Consider the plot below of \( h_{\theta}(x) = \theta_0 + \theta_1 x \). What are \( \theta_0 \) and \( \theta_1 \)?

- \( \theta_0 = 0, \theta_1 = 1 \)
- \( \theta_0 = 0.5, \theta_1 = 1 \)
- \( \theta_0 = 1, \theta_1 = 0.5 \)
- \( \theta_0 = 1, \theta_1 = 1 \)
Idea: Choose $\theta_0, \theta_1$ so that $h_\theta(x)$ is close to $y$ for our training examples $(x, y)$.

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

Cost function

Squared error function

$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$
Linear regression with one variable

Cost function intuition

Machine Learning
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: minimize \( J(\theta_0, \theta_1) \)

Simplified:

\[
h_\theta(x) = \theta_1 x
\]

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

Goal: minimize \( J(\theta_1) \)
\[ h_\theta(x) \]

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

\[ J(\theta_1) \]

(function of the parameter \( \theta_1 \))

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

\[ \theta_1 = 0.5 \]

\[ J(0.5) = 0 \]
\[ h_{\theta}(x) \]
(for fixed \( \theta_1 \), this is a function of \( x \))

\[ J(\theta_1) \]
(function of the parameter \( \theta_1 \))

\[
J(0.5) = \frac{1}{2m} \left[ (0.5-1)^2 + (1-2)^2 + (1.5-3)^2 \right]
= \frac{1}{2 \times 3} \times 3.5 = \frac{3.5}{6} \approx 0.58
\]

\( \theta_1 = 0.5 ? \)
\( J(0) = ? \)
Suppose we have a training set with $m=3$ examples, plotted below. Our hypothesis representation is $h_\theta(x) = \theta_1 x$, with parameter $\theta_1$. The cost function $J(\theta_1)$ is $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$. What is $J(0)$?

- $0$
- $1/6$
- $1$
- $14/6$
Suppose we have a training set with $m=3$ examples, plotted below. Our hypothesis representation is $h_\theta(x) = \theta_1 x$, with parameter $\theta_1$. The cost function $J(\theta_1)$ is $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$. What is $J(0)$?
\[ h_\theta(x) \]
(for fixed \( \theta_1 \), this is a function of \( x \))

\[ J(0) = \frac{1}{2m} (1^2 + 2^2 + 3^2) \]
\[ = \frac{1}{6} \cdot 14 \approx 2.3 \]

\[ h(x) = -0.5 \]

\[ J(\theta_1) \]
(function of the parameter \( \theta_1 \))
Linear regression with one variable

Cost function intuition II

Machine Learning
Hypothesis: $h_{\theta}(x) = \theta_0 + \theta_1 x$

Parameters: $\theta_0, \theta_1$

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

Goal: minimize $J(\theta_0, \theta_1)$
\( h_\theta(x) \)

(for fixed \( \theta_0, \theta_1 \), this is a function of \( x \))

\[
h_\theta(x) = 50 + 0.06x
\]

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

(function of the parameters \( \theta_0, \theta_1 \))

Price ($)
in 1000’s

Size in feet\(^2\) (\( x \))

\( \theta_0 = 50 \)

\( \theta_1 = 0.06 \)
\[ h_\theta(x) \]

(for fixed \( \theta_0, \theta_1 \), this is a function of \( x \))

\[ J(\theta_0, \theta_1) \]

(function of the parameters \( \theta_0, \theta_1 \))
\( h_\theta(x) \)

(for fixed \( \theta_0, \theta_1 \), this is a function of \( x \))

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

(function of the parameters \( \theta_0, \theta_1 \))
\( h_\theta(x) \)

(for fixed \( \theta_0, \theta_1 \), this is a function of \( x \))

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

(function of the parameters \( \theta_0, \theta_1 \))
\[ h_\theta(x) \]

(for fixed \( \theta_0, \theta_1 \), this is a function of \( x \))

\[ J(\theta_0, \theta_1) \]

(function of the parameters \( \theta_0, \theta_1 \))
Linear regression with one variable

Next Seminar

Gradient descent

Machine Learning