Machine Learning is...

Machine learning is about predicting the future based on the past.

-- Hal Daume III
Machine Learning is...

Machine learning is about predicting the future based on the past.

-- Hal Daume III
Data

examples

Data

apple

banana

apple

banana
Data

examples

Data

IMDb

IMDb

IMDb

IMDb
Data

examples
Data

examples

Data
Supervised learning: given labeled examples.
Supervised learning: given labeled examples
Supervised learning: learn to predict new example.
Supervised learning: classification

- apple
- banana

Classification: a finite set of labels

Supervised learning: given labeled examples
Classification Example
Classification Applications

Face recognition

Character recognition

Spam detection

Medical diagnosis: From symptoms to illnesses

Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
Supervised learning: regression

Regression: label is real-valued

Supervised learning: given labeled examples
Regression Example

Price of a used car

\( x \): car attributes (e.g. mileage)

\( y \): price
Regression Applications

Economics/Finance: predict the value of a stock

Epidemiology

Car/plane navigation: angle of the steering wheel, acceleration, ...

Temporal trends: weather over time
Supervised learning: given labeled examples

Ranking: label is a ranking
Ranking example

Given a query and a set of web pages, rank them according to relevance.
Ranking Applications

User preference, e.g. Netflix “My List” -- movie queue ranking

iTunes

flight search (search in general)

reranking N-best output lists
Unsupervised learning

Unsupervised learning: given data, i.e. examples, but no labels
Unsupervised learning applications

learn clusters/groups without any label

customer segmentation (i.e. grouping)

image compression

bioinformatics: learn motifs
Given a *sequence* of examples/states and a *reward* after completing that sequence, learn to predict the action to take in for an individual example/state.
Reinforcement learning example

Backgammon

Given sequences of moves and whether or not the player won at the end, learn to make good moves
Other learning variations

What data is available:
- Supervised, unsupervised, reinforcement learning
- Semi-supervised, active learning, ...

How are we getting the data:
- Online vs. offline learning

Type of model:
- Generative vs. discriminative
- Parametric vs. non-parametric
Representing examples

What is an example?
How is it represented?
Features

Examples: $f_1, f_2, f_3, \ldots, f_n$

Features: $f_1, f_2, f_3, \ldots, f_n$

How our algorithms actually “view” the data

Features are the questions we can ask about the examples
Features

How our algorithms actually “view” the data

Features are the questions we can ask about the examples

examples

features

red, round, leaf, 3oz, ...

green, round, no leaf, 4oz, ...

yellow, curved, no leaf, 8oz, ...

green, curved, no leaf, 7oz, ...
During learning/training/induction, learn a model of what distinguishes apples and bananas based on the features.
Classification revisited

The model can then classify a new example based on the features.

red, round, no leaf, 4oz, ...

model/classifier

predict

Apple or banana?
Classification revisited

The model can then classify a new example based on the features.
## Classification revisited

<table>
<thead>
<tr>
<th>Training data</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>examples</strong></td>
<td><strong>label</strong></td>
</tr>
<tr>
<td>red, round, leaf, 3oz, ...</td>
<td>apple</td>
</tr>
<tr>
<td>green, round, no leaf, 4oz, ...</td>
<td>apple</td>
</tr>
<tr>
<td>yellow, curved, no leaf, 4oz, ...</td>
<td>banana</td>
</tr>
<tr>
<td>green, curved, no leaf, 5oz, ...</td>
<td>banana</td>
</tr>
</tbody>
</table>
Classification revisited

<table>
<thead>
<tr>
<th>Training data</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>examples</td>
<td>label</td>
</tr>
<tr>
<td>red, round, leaf, 3oz, ...</td>
<td>apple</td>
</tr>
<tr>
<td>green, round, no leaf, 4oz, ...</td>
<td>apple</td>
</tr>
<tr>
<td>yellow, curved, no leaf, 4oz, ...</td>
<td>banana</td>
</tr>
<tr>
<td>green, curved, no leaf, 5oz, ...</td>
<td>banana</td>
</tr>
</tbody>
</table>

Learning is about **generalizing** from the training data.
A simple machine learning example

http://www.mindreaderpro.appspot.com/
models

We have many, many different options for the model.

They have different characteristics and perform differently (accuracy, speed, etc.)
Probabilistic modeling

Model the data with a probabilistic model which tells us how likely a given data example is
Probabilistic models

Example to label
yellow, curved, no leaf, 6oz
features

probabilistic model:
p(example)

apple or banana
Probabilistic models

For each label, ask for the probability

features

label

probabilistic model:

\( p(\text{example}) \)

yellow, curved, no leaf, 6oz, banana

yellow, curved, no leaf, 6oz, apple
Probabilistic models

Pick the label with the highest probability

yellow, curved, no leaf, 6oz, banana

p(example) = 0.004

yellow, curved, no leaf, 6oz, apple

p(example) = 0.00002

features → label → probabilistic model
Probability basics

A probability distribution gives the probabilities of all possible values of an event.

For example, say we flip a coin three times. We can define the probability of the number of times the coin came up heads.

<table>
<thead>
<tr>
<th>$P(\text{num heads})$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(3) = ?$</td>
<td></td>
</tr>
<tr>
<td>$P(2) = ?$</td>
<td></td>
</tr>
<tr>
<td>$P(1) = ?$</td>
<td></td>
</tr>
<tr>
<td>$P(0) = ?$</td>
<td></td>
</tr>
</tbody>
</table>
What are the possible outcomes of three flips (hint, there are eight of them)?

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H
Assuming the coin is fair, what are our probabilities?

\[
\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}
\]

|     | T | T | T | T | T | H | T | T | T | H | H | T | T | T | H | T | T | H | H | T | T | H | H | T | H | H | H |
| P(3)|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| P(2)|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| P(1)|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| P(0)|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Probability distributions

Assuming the coin is fair, what are our probabilities?

probability = \frac{\text{number of times it happens}}{\text{total number of cases}}

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

<table>
<thead>
<tr>
<th>P(num heads)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(3) = ?</td>
<td></td>
</tr>
<tr>
<td>P(2) = ?</td>
<td></td>
</tr>
<tr>
<td>P(1) = ?</td>
<td></td>
</tr>
<tr>
<td>P(0) = ?</td>
<td></td>
</tr>
</tbody>
</table>
Probability distributions

Assuming the coin is fair, what are our probabilities?

probability = \frac{\text{number of times it happens}}{\text{total number of cases}}

<table>
<thead>
<tr>
<th>Outcome</th>
<th>P(num heads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T T T</td>
<td>P(3) = 1/8</td>
</tr>
<tr>
<td>T T H</td>
<td>P(2) = ?</td>
</tr>
<tr>
<td>T H T</td>
<td>P(1) = ?</td>
</tr>
<tr>
<td>T H H</td>
<td>P(0) = ?</td>
</tr>
<tr>
<td>H T T</td>
<td></td>
</tr>
<tr>
<td>H T H</td>
<td></td>
</tr>
<tr>
<td>H H T</td>
<td></td>
</tr>
<tr>
<td>H H H</td>
<td></td>
</tr>
</tbody>
</table>
Assuming the coin is fair, what are our probabilities?

Probability = \frac{\text{number of times it happens}}{\text{total number of cases}}

<table>
<thead>
<tr>
<th>T T T T</th>
<th>\text{P(num heads)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>T T H H</td>
<td>P(3) = 1/8</td>
</tr>
<tr>
<td>T H T T</td>
<td>P(2) = ?</td>
</tr>
<tr>
<td>T H H H</td>
<td>P(1) = ?</td>
</tr>
<tr>
<td>H T T T</td>
<td>P(0) = ?</td>
</tr>
<tr>
<td>H T H H</td>
<td></td>
</tr>
<tr>
<td>H H T T</td>
<td></td>
</tr>
<tr>
<td>H H H H</td>
<td></td>
</tr>
</tbody>
</table>
Probability distributions

Assuming the coin is fair, what are our probabilities?

\[
p\text{\textunderscore probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}
\]

<table>
<thead>
<tr>
<th>T T T T</th>
<th>P(num heads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T T T H</td>
<td>P(3) = 1/8</td>
</tr>
<tr>
<td>T H T T</td>
<td>P(2) = 3/8</td>
</tr>
<tr>
<td>T H H H</td>
<td>P(1) = ?</td>
</tr>
<tr>
<td>H T T T</td>
<td>P(0) = ?</td>
</tr>
<tr>
<td>H T H H</td>
<td></td>
</tr>
<tr>
<td>H H T T</td>
<td></td>
</tr>
<tr>
<td>H H H H</td>
<td></td>
</tr>
</tbody>
</table>
Probability distributions

Assuming the coin is fair, what are our probabilities?

Probability = \[
\frac{\text{number of times it happens}}{\text{total number of cases}}\]

\[
\begin{array}{c}
\text{T T T} \\
\text{T T H} \\
\text{T H T} \\
\text{T H H} \\
\text{H T T} \\
\text{H T H} \\
\text{H H T} \\
\text{H H H}
\end{array}
\]

\[
\begin{array}{|l|}
\hline
\text{P(num heads)} & \\
\hline
\text{P(3) = 1/8} & \\
\text{P(2) = 3/8} & \\
\text{P(1) = 3/8} & \\
\text{P(0) = 1/8} & \\
\hline
\end{array}
\]
A probability distribution assigns probability values to all possible values.

Probabilities are between 0 and 1, inclusive.

The sum of all probabilities in a distribution must be 1.

<table>
<thead>
<tr>
<th>P(num heads)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(3)</td>
<td>1/8</td>
</tr>
<tr>
<td>P(2)</td>
<td>3/8</td>
</tr>
<tr>
<td>P(1)</td>
<td>3/8</td>
</tr>
<tr>
<td>P(0)</td>
<td>1/8</td>
</tr>
</tbody>
</table>
Probability distribution

A probability distribution assigns probability values to all possible values.

Probabilities are between 0 and 1, inclusive.

The sum of all probabilities in a distribution must be 1.
Some example probability distributions

probability of heads
(distribution options: heads, tails)

probability of passing class
(distribution options: pass, fail)

probability of rain today
(distribution options: rain or no rain)

probability of getting an ‘A’
(distribution options: A, B, C, D, F)
Conditional probability distributions

Sometimes we may know extra information about the world that may change our probability distribution.

$P(X|Y)$ captures this (read “probability of $X$ given $Y$”)

- Given some information ($Y$) what does our probability distribution look like?
- Note that this is still just a normal probability distribution.
Conditional probability example

<table>
<thead>
<tr>
<th>P(pass 51a)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(pass)</td>
<td>0.9</td>
</tr>
<tr>
<td>P(not pass)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Unconditional probability distribution
Conditional probability example

\[ P(\text{pass 51a}) \]
\[ P(\text{pass}) = 0.9 \]
\[ P(\text{not pass}) = 0.1 \]

\[ P(\text{pass 51a | don’t study}) \]
\[ P(\text{pass}) = 0.5 \]
\[ P(\text{not pass}) = 0.5 \]

\[ P(\text{pass 51a | do study}) \]
\[ P(\text{pass}) = 0.95 \]
\[ P(\text{not pass}) = 0.05 \]
Conditional probability example

<table>
<thead>
<tr>
<th>$P(\text{rain in LA})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\text{rain}) = 0.05$</td>
</tr>
<tr>
<td>$P(\text{no rain}) = 0.95$</td>
</tr>
</tbody>
</table>

Unconditional probability distribution
Conditional probability example

P(rain in LA| January)
- P(rain) = 0.2
- P(no rain) = 0.8

P(rain in LA| not January)
- P(pass) = 0.03
- P(not pass) = 0.97

P(rain in LA)
- P(rain) = 0.05
- P(no rain) = 0.95

Still probability distributions over passing or rain in LA

Conditional probability distributions
Joint distribution

Probability over two events: $P(X,Y)$

Has probabilities for all possible combinations over the two events

<table>
<thead>
<tr>
<th>51Pass, EngPass</th>
<th>$P(51\text{Pass}, \text{EngPass})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true, true</td>
<td>.88</td>
</tr>
<tr>
<td>true, false</td>
<td>.01</td>
</tr>
<tr>
<td>false, true</td>
<td>.04</td>
</tr>
<tr>
<td>false, false</td>
<td>.07</td>
</tr>
</tbody>
</table>
Joint distribution

Still a probability distribution

**All** questions/probabilities that we might want to ask about these two things can be calculated from the joint distribution

<table>
<thead>
<tr>
<th>51Pass, EngPass</th>
<th>P(51Pass, EngPass)</th>
</tr>
</thead>
<tbody>
<tr>
<td>true, true</td>
<td>.88</td>
</tr>
<tr>
<td>true, false</td>
<td>.01</td>
</tr>
<tr>
<td>false, true</td>
<td>.04</td>
</tr>
<tr>
<td>false, false</td>
<td>.07</td>
</tr>
</tbody>
</table>

What is P(51pass = true)?
Joint distribution

<table>
<thead>
<tr>
<th>51Pass, EngPass</th>
<th>P(51Pass, EngPass)</th>
</tr>
</thead>
<tbody>
<tr>
<td>true, true</td>
<td>.88</td>
</tr>
<tr>
<td>true, false</td>
<td>.01</td>
</tr>
<tr>
<td>false, true</td>
<td>.04</td>
</tr>
<tr>
<td>false, false</td>
<td>.07</td>
</tr>
</tbody>
</table>

There are two ways that a person can pass 51: they can do it while passing or not passing English.

\[
P(51\text{Pass}=\text{true}) = P(\text{true, true}) + P(\text{true, false}) = 0.89
\]
Relationship between distributions

\[ P(X, Y) = P(Y) \times P(X|Y) \]

Can think of it as describing the two events happening in two steps:

The likelihood of X and Y happening:
1. How likely it is that Y happened?
2. Given that Y happened, how likely is it that X happened?
Relationship between distributions

\[ P(51\text{Pass}, \text{EngPass}) = P(\text{EngPass}) \times P(51\text{Pass}|\text{EngPass}) \]

The probability of passing CS51 and English is:
1. Probability of passing English *
2. Probability of passing CS51 given that you passed English
Relationship between distributions

\[ P(\text{51Pass}, \text{EngPass}) = P(\text{51Pass}) \times P(\text{EngPass|51Pass}) \]

The probability of passing CS51 and English is:
1. Probability of passing CS51 *
2. Probability of passing English \textbf{given} that you passed CS51

Can also view it with the other event happening first