Lecture 1
Machine Learning Basics

Sung Kim <hunkim+ml@gmail.com>
Basic concepts

• What is ML?
• What is learning?
  - supervised
  - unsupervised
• What is regression?
• What is classification?
Machine Learning

• Limitations of explicit programming
  - Spam filter: many rules
  - Automatic driving: too many rules

• Machine learning: "Field of study that gives computers the ability to learn without being explicitly programmed" Arthur Samuel (1959)
Supervised/Unsupervised learning

- Supervised learning:
  - learning with labeled examples - training set
Supervised learning

An example training set for four visual categories.

http://cs231n.github.io/classification/
Supervised/Unsupervised learning

- **Supervised learning:** learning with labeled examples
- **Unsupervised learning:** un-labeled data
  - Google news grouping
  - Word clustering
Supervised learning

- Most common problem type in ML
  - **Image labeling**: learning from tagged images
  - **Email spam filter**: learning from labeled (spam or ham) email
  - **Predicting exam score**: learning from previous exam score and time spent
Training data set
AlphaGo
Types of supervised learning

• Predicting final exam score based on time spent
  - regression

• Pass/non-pass based on time spent
  - binary classification

• Letter grade (A, B, C, E and F) based on time spent
  - multi-label classification
Predicting final exam score based on time spent

<table>
<thead>
<tr>
<th>x (hours)</th>
<th>y (score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>
Pass/non-pass based on time spent

<table>
<thead>
<tr>
<th>x (hours)</th>
<th>y (pass/fail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>P</td>
</tr>
<tr>
<td>9</td>
<td>P</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
</tr>
</tbody>
</table>
Letter grade (A, B, ...) based on time spent

<table>
<thead>
<tr>
<th>x (hours)</th>
<th>y (grade)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
</tr>
</tbody>
</table>
Next
Linear regression