Introduction to Machine Learning (and Notation)

David I. Inouye
Saturday, August 24, 2019
Announcements

- TA: Liming Wu

- Homework 1 will be posted by Wed due next Wed
  - Submit GitHub username ASAP: https://forms.gle/A4to4Q7huAiKaQBN9

- Hopefully, first quiz on Wednesday, beginning of class
Outline

▸ Supervised learning
  ▸ Regression
  ▸ Classification

▸ Unsupervised learning

▸ Other key concepts
The goal of **supervised learning** is to estimate a **mapping (or function)** between input and output.
The goal of **supervised learning** is to estimate a mapping (or function) between input and output given only input-output examples.
The set of input-output pairs is called a training set, denoted by \( D = \{(x_i, y_i)\}_{i=1}^{n} \)

- **Input** \( x_i \)
  - Called features (ML), attributes, or covariates (Stats). Sometimes just variables.
  - Can be numeric, categorical, discrete, or nominal.
  - Examples
    - \([\text{height, weight, age, gender}]\)
    - \([x_1, x_2, \ldots, x_d]\) – A \(d\)-dimensional vector of numbers
    - Image
    - Email message

- **Output** \( y_i \)
  - Called output, response, or target (or label)
  - Real-valued/numeric output: e.g., \( y_i \in \mathbb{R} \)
  - Categorical, discrete, or nominal output: \( y_i \) from finite set, i.e., \( y_i \in \{1,2,\ldots,c\} \)
If the output $y_i$ is numeric, then the problem is known as regression.

NOTE: Input $x$ does not have to be numeric. Only the output $y$ must be numeric.
If the output $y_i$ is numeric, then the problem is known as **regression**

- Given height $x_i$, predict age $y_i$
- Predict GPA given SAT score
- Predict SAT score given GPA
- Predict GRE given SAT and GPA
If output is categorical, then the problem is known as classification.

- Predicted: cat
- Predicted: dog
- Predicted: cat
- Predicted: dog
If output is categorical, then the problem is known as classification

- Given height $x$, predict “male” ($y = 0$) or “female” ($y = 1$)

- Predict defaulting on loan (“yes” or “no”) given salary and mortgage payment
Side note: **Encoding / representing** a categorical variable can be done in many ways

▶ Suppose the categorical variable is “yes” and “no”
  ▶ Canonical ways: “no” -> 0 and “yes” -> 1
  ▶ What are other possible encodings?

▶ What if there are more than two categories such as cats, dogs, fish and snakes?

▶ What is good and bad about using \{1,2,3,4\} for above example of animals?

▶ One-hot encoding is another common way
The goal of unsupervised learning is to find “interesting patterns” ONLY in the input

- Also called descriptive learning or knowledge discovery

- What are “interesting patterns”?  
  - Could be many things  
  - Clusters  
  - Correlations
In unsupervised learning, the training set is only a set of input values $\mathcal{D} = \{x_i\}_{i=1}^n$

- Estimate natural clusters (or groups) of customers

- Estimate the correlation between height and weight, $x = [h, w]$

- Estimate a single number that summarizes all variables of wealth (e.g. credit score)
Given this dataset, should we use supervised or unsupervised learning?

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.
Is this a regression or classification problem?

\[ d \text{ features/attributes/covariates} \]

\[ n \text{ samples/observations/examples} \]

<table>
<thead>
<tr>
<th>Color</th>
<th>Shape</th>
<th>Size (cm)</th>
<th>Is it good?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Square</td>
<td>10</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>2.4</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>20.7</td>
<td>no</td>
</tr>
</tbody>
</table>

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.
Suppose we assume classification, which features are the input $x$ and which are the output $y$?

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.
Suppose we assume regression, which features are the input $\mathbf{x}$ and which are the output $\mathbf{y}$?

$d$ features/attributes/covariates

<table>
<thead>
<tr>
<th>Color</th>
<th>Shape</th>
<th>Size (cm)</th>
<th>Is it good?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Square</td>
<td>10</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>2.4</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>20.7</td>
<td>no</td>
</tr>
</tbody>
</table>

$n$ samples/observations/examples

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.
How could we use unsupervised learning?

$\hat{\mathbf{x}}_{\hat{n}}$ samples/observations/examples

$d$ features/attributes/covariates

<table>
<thead>
<tr>
<th>Color</th>
<th>Shape</th>
<th>Size (cm)</th>
<th>Is it good?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Square</td>
<td>10</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>2.4</td>
<td>yes</td>
</tr>
<tr>
<td>Red</td>
<td>Ellipse</td>
<td>20.7</td>
<td>no</td>
</tr>
</tbody>
</table>

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.
The dataset cannot determine the task, rather the context determines the task.

Adapted from Machine Learning: A Probabilistic Perspective, Ch. 1, Kevin P. Murphy, 2012.