Instance-Based Learning

CS4780/5780 – Machine Learning
Fall 2014
Thorsten Joachims
Cornell University

Reading: Mitchell Chapter 1 & Sections 8.1 - 8.2
Optional Reading: Linden et al., Amazon Recommendations (http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf)

Concept Learning

- **Definition:**
  Acquire an operational definition of a general category of objects given positive and negative training examples.

  Also called: binary classification, binary supervised learning...

Concept Learning Example

- **Instance Space X:** Set of all possible objects describable by attributes (often called features).
- **Concept c:** Subset of objects from X (c is unknown).
- **Target Function f:** Characteristic function indicating membership in c based on attributes (i.e. label) (f is unknown).
- **Example (x,y):** Instance x with label y=f(x).
- **Training Data S:** Collection of examples observed by learning algorithm.

K-Nearest Neighbor (KNN)

- **Given:** Training data \((\vec{x}_1, y_1), ..., (\vec{x}_n, y_n)\)
  - Attribute vectors: \(\vec{x} \in X\)
  - Labels: \(y \in Y\)
- **Parameter:**
  - Similarity function: \(K : X \times X \to \mathbb{R}\)
  - Number of nearest neighbors to consider: \(k\)
- **Prediction rule**
  - New example \(x'\)
  - K-nearest neighbors: \(k\) train examples with largest \(K(\vec{x}, \vec{x}')\)

\[
h(x') = \arg\max_{y \in Y} \left\{ \sum_{i \in knn(x')} 1_{y_i = y} \right\}
\]

KNN Example

- **How will new examples be classified?**
  - Similarity function?
  - Value of \(k\)?

\[
h(x') = \arg\max_{y \in Y} \left\{ \sum_{i \in knn(x')} 1_{y_i = y} \right\}
\]
Weighted K-Nearest Neighbor

• Given: Training data \((x_1, y_1), \ldots, (x_n, y_n)\)
  - Attribute vectors: \(x_i \in X\)
  - Target attribute: \(y_i \in Y\)
• Parameter:
  - Similarity function: \(K : X \times X \rightarrow \mathbb{R}\)
  - Number of nearest neighbors to consider: \(k\)
• Prediction rule
  - New example \(x'\)
  - K-nearest neighbors: \(k\) train examples with largest \(K(x_i, x')\)

\[
h(x') = \arg \max_{y' \in Y} \left\{ \sum_{i \in \text{knn}(x')} 1[y_i = y] K(x_i, x') \right\}
\]

Supervised Learning

• Task:
  - Learn (to imitate) a function \(f : X \rightarrow Y\)
• Training Examples:
  - Learning algorithm is given the correct value of the function for particular inputs \(\rightarrow\) training examples
  - An example is a pair \((x, f(x))\), where \(x\) is the input and \(f(x)\) is the output of the function applied to \(x\).
• Goal:
  - Find a function \(h : X \rightarrow Y\)
    - that approximates \(f : X \rightarrow Y\)
    - as well as possible.

Types of Attributes

• Symbolic (nominal)
  - EyeColor (brown, blue, green)
• Boolean
  - alive (TRUE,FALSE)
• Numeric
  - Integer: age \([0, 105]\)
  - Real: height
• Structural
  - Natural language sentence: parse tree
  - Protein: sequence of amino acids

Example: Expensive Housing (> $200 / sqft)

Example: Effect of k

Weighted K-NN for Regression

• Given: Training data \((x_1, y_1), \ldots, (x_n, y_n)\)
  - Attribute vectors: \(x_i \in X\)
  - Target attribute: \(y_i \in \mathbb{R}\)
• Parameter:
  - Similarity function: \(K : X \times X \rightarrow \mathbb{R}\)
  - Number of nearest neighbors to consider: \(k\)
• Prediction rule
  - New example \(x'\)
  - K-nearest neighbors: \(k\) train examples with largest \(K(x_i, x')\)

\[
h(x') = \frac{\sum_{i \in \text{knn}(x')} y_i K(x_i, x')}{\sum_{i \in \text{knn}(x')} K(x_i, x')}
\]
Collaborative Filtering

<table>
<thead>
<tr>
<th>Rating</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>$m_4$</th>
<th>$m_5$</th>
<th>$m_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u_2$</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$u_3$</td>
<td>2</td>
<td>4</td>
<td></td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>$u$</td>
<td>?</td>
<td>1</td>
<td>4</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

"The images of movies are not clear."

"Collaborative Filtering is a method in machine learning and data mining, where an individual's data is predicted and improved by the similarity of related data."

"It is often used in recommendation systems, where an individual's preferences are predicted and recommended content is generated based on similar patterns observed in the data of others."

"The rating matrix above represents the ratings given by different users for various movies. The columns represent different movies, and the rows represent different users. The entries indicate the ratings given by each user for each movie."

"The '?' entries indicate missing or unobserved ratings. Collab filtering can predict these missing ratings based on the similarities of ratings from the observed data."

"In this case, the system might predict a rating for the movie 'u' based on the patterns observed in the ratings of 'u_1' and 'u_2' and 'u_3'."