Multi-Class Classification

Dan Sheldon

October 8, 2014
A Real Classification Problem

Classify handwritten digits.

\[ y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \]

We don’t know how to solve this yet
Hand-written digit classification

Input: 20 × 20 grayscale image

Unroll the image into a vector

\[
\begin{bmatrix}
  x_1 & x_{21} & \cdots & x_{381} \\
  x_2 & x_{22} & \cdots & x_{382} \\
  \vdots \\
  x_{20} & x_{40} & \cdots & x_{400}
\end{bmatrix}
\]

Feature vector \( \mathbf{x} \in \mathbb{R}^{400} \)

\[
\mathbf{x} = (x_1, \ldots, x_{400})^T
\]
Multi-class Classification

Input: $\mathbf{x} \in \mathbb{R}^m$ (continuous or discrete)
Labels: $y \in \{1, \ldots, K\}$

Exercise: solve using logistic regression

- Use one or more binary ($y \in \{0, 1\}$) classifiers
- Hint: think about prediction first, then training.
Visualization

Format weight vector as an image:

Recall that

\[
\text{Prediction} = \begin{cases} 
1 & \theta^T x \geq 0 \\
0 & \theta^T x < 0 
\end{cases}
\]

Dot product = multiply together corresponding pixels and add
One vs. All Classification

Learn a separate classifier for each class $c = 1, \ldots, K$

Let $y_c^{(i)} = \begin{cases} 1 & \text{if } y^{(i)} = c \\ 0 & \text{otherwise} \end{cases}$

<table>
<thead>
<tr>
<th>$x^T$</th>
<th>$y$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Training:** for each class $c$, fit a binary classifier using training labels $y_c^{(i)}$ to get parameter vector $\theta_c$

**Prediction:** make a prediction for each class and choose the one with highest probability

\[ \text{predict } y = \arg\max_c h_{\theta_c}(x) \]
Visualization: One vs. All
One vs. One

Fit a classifier for each pair of classes

Let $y_{cd}^{(i)} = \begin{cases} 1 & \text{if } y^{(i)} = c \\ 0 & \text{if } y^{(i)} = d \end{cases}$

<table>
<thead>
<tr>
<th>$x^T$</th>
<th>$y$</th>
<th>$y_{12}$</th>
<th>$y_{13}$</th>
<th>$y_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Training: for each pair $c \neq d$, fit a binary classifier with labels $y_{cd}^{(i)}$

using only examples from class $c$ or $d$

▶ Result: parameter vector $\theta_{cd}$

Prediction: voting scheme. Explain on board.
Visualization: One vs. One
Learn a separate classifier for each bit of codeword

<table>
<thead>
<tr>
<th>Class</th>
<th>$f_0$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$f_3$</th>
<th>$f_4$</th>
<th>$f_5$</th>
<th>$f_6$</th>
<th>$f_7$</th>
<th>$f_8$</th>
<th>$f_9$</th>
<th>$f_{10}$</th>
<th>$f_{11}$</th>
<th>$f_{12}$</th>
<th>$f_{13}$</th>
<th>$f_{14}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Dietterich and Bakiri 1995
(Possible project idea)