Quiz 4 Solutions

1. Suppose you have trained three classifiers, $h_1$, $h_2$, and $h_3$, returning binary classifications (+1 or -1). The observed accuracy of these classifiers are as follows: $h_1: 51\%$, $h_2: 53\%$, $h_3: 55\%$.

Assuming that the errors of these classifiers are independent, what is the predicted accuracy of an ensemble hypothesis $H$, where $H$'s classification of an instance is a majority vote of the classifications of $h_1$, $h_2$, and $h_3$?

**SOLUTION:**

Let $C$ denote "correct". Below are listed only the cases in which $H$ is correct.

<table>
<thead>
<tr>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$h_3$</th>
<th>$H$</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>(.51)(.53)(.55) = .149</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>I</td>
<td>C</td>
<td>(.51)(.53)(.45) = .122</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>C</td>
<td>C</td>
<td>(.51)(.47)(.55) = .132</td>
</tr>
<tr>
<td>I</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>(.49)(.53)(.55) = .142</td>
</tr>
</tbody>
</table>

Total probability that $H$ is correct is .545
Useful formulas for Adaboost:

\[ \epsilon_i = \sum_{j=1}^{N} w_j(j) \delta(y_j \neq h_i(x_j)), \text{ where} \]

\[ \delta(y_j \neq h_i(x_j)) = \begin{cases} 
1 & \text{if } y_j \neq h_i(x_j) \\
0 & \text{otherwise} 
\end{cases} \]

\[ \alpha_i = \frac{1}{2} \ln \left( \frac{1 - \epsilon_i}{\epsilon_i} \right) \]

\[ w_{t+1}(i) = \frac{w_t(i) \exp(-\alpha_i y_i h_t(x_i))}{Z_t} \]

2. Suppose you have done one iteration of Adaboost to produce classifier \( h_1 \) and find that \( h_1 \) has the following results on the training data. (Assume the initial weights on the training data are uniform.)

<table>
<thead>
<tr>
<th>Example</th>
<th>True Class</th>
<th>Predicted Class (( h_1(x) ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

(a) What is \( \epsilon_1 \) ?

**SOLUTION:**

\[ \epsilon_1 = .25 \]

(b) What is \( \alpha_1 \) ?

**SOLUTION:**

\[ \alpha_1 = \frac{1}{2} \ln \frac{.75}{.25} = .55 \]
(c) What is the new probability distribution on the training set?

**SOLUTION:**

\[ \hat{w}_2(1) = 0.25 \exp(-0.55(1)(1)) = 0.144 \]

\[ \hat{w}_2(2) = 0.25 \exp(-0.55(-1)(-1)) = 0.144 \]

\[ \hat{w}_2(3) = 0.25 \exp(-0.55(-1)(1)) = 0.433 \]

\[ \hat{w}_2(4) = 0.25 \exp(-0.55(1)(1)) = 0.144 \]

\[ Z_1 = 0.865 \]

\[ w_2(1) = \frac{0.144}{0.865} = 0.166 \]

\[ w_2(2) = \frac{0.144}{0.865} = 0.166 \]

\[ w_2(3) = \frac{0.433}{0.865} = 0.501 \]

\[ w_2(4) = \frac{0.144}{0.865} = 0.166 \]
(d) Suppose you have finished running Adaboost to produce classifiers $h_1$ and $h_2$, with $\alpha_1 = .2$ and $\alpha_2 = -.4$. You run $h_1$ and $h_2$ on a test instance $x$ and get the following results:

<table>
<thead>
<tr>
<th>Example</th>
<th>True Class</th>
<th>Predicted Class ($h_1(x)$)</th>
<th>Predicted Class ($h_2(x)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>+1</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
</tbody>
</table>

What is the class predicted by the ensemble classifier $H(x)$, using $h_1$ and $h_2$?

**SOLUTION:**

$H(x) = \text{sgn} (.2 h_1(x) - .4 h_2(x)) = \text{sgn} (-.2 + .4) = +1$

Class is positive.

3. Let $S = (x_1, x_2, x_3, x_4)$, where $x_1 = (0, 2), x_2 = (0, 3), x_3 = (0, 0), x_4 = (8, 1)$.

Suppose you are given two initial cluster centers: $m_1 = (0, 0), m_2 = (2, 1)$

Showing your work, simulate by hand **two iterations** of K-means clustering. In each iteration: (1) Assign data points to a cluster center. (2) Update the cluster centers.

(Hint: for ease of calculation, you can simply calculate $d^2$, where $d =$ Euclidean distance, rather than $d$ itself).

**SOLUTION:**

**Iteration 1:**

$d^2(x_1, m_1) = 0^2 + 2^2 = 4$ \quad $d^2(x_1, m_2) = 2^2 + 1^2 = 5$

$d^2(x_2, m_1) = 0^2 + 3^2 = 9$ \quad $d^2(x_2, m_2) = 2^2 + 2^2 = 8$

$d^2(x_3, m_1) = 0^2 + 0^2 = 0$ \quad $d^2(x_3, m_2) = 2^2 + 1^2 = 5$

$d^2(x_4, m_1) = 8^2 + 1^2 = 65$ \quad $d^2(x_4, m_2) = 6^2 + 0^2 = 36$

**Cluster membership:**

$m_1: x_1, x_3$ \quad $m_2: x_2, x_4$

**Updated cluster centers:**

$m_1: \frac{(0.2) + (0.0)}{2} = (0.1)$ \quad $m_2: \frac{(0.3) + (8.1)}{2} = \frac{(8.4)}{2} = (4.2)$
Iteration 2:

\[ d^2(x_1, m_1) = 0^2 + 1^2 = 1 \quad d^2(x_1, m_2) = 4^2 + 0^2 = 16 \]
\[ d^2(x_2, m_1) = 0^2 + 2^2 = 4 \quad d^2(x_2, m_2) = 4^2 + 1^2 = 17 \]
\[ d^2(x_3, m_1) = 0^2 + 1^2 = 1 \quad d^2(x_3, m_2) = 4^2 + 2^2 = 20 \]
\[ d^2(x_4, m_1) = 8^2 + 0^2 = 64 \quad d^2(x_4, m_2) = 4^2 + 1^2 = 17 \]

Cluster membership:

\[ m_1: x_1, x_2, x_3 \]
\[ m_2: x_4 \]

Updated cluster centers:

\[ m_1: \frac{(0,2)+(0,3)+(0,0)}{3} = (0, \frac{5}{3}) \]
\[ m_2: (8,1) \]

4. List three potential weaknesses of (or issues for) K-means clustering.

**SOLUTION:** See slides for list of potential weaknesses.