Quiz 4

Please write all answers on these pages.

1. Consider the following training data, where each instance \( x \) is described by two features, \( F_1 \) and \( F_2 \). The possible classes are POS and NEG.

<table>
<thead>
<tr>
<th>Instance</th>
<th>( F_1 )</th>
<th>( F_2 )</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>4</td>
<td>4</td>
<td>POS</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>2</td>
<td>4</td>
<td>POS</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>4</td>
<td>2</td>
<td>NEG</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>6</td>
<td>2</td>
<td>NEG</td>
</tr>
</tbody>
</table>

Give the probabilistic model that Gaussian Naive Bayes would compute—that is, the parameters of the relevant Gaussian distributions.

**Hint:** Recall that

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x - \mu)^2}{n}}
\]

**SOLUTION:**

**Probabilistic model:**

\( P(\text{POS}) = .5 \)

\( P(\text{NEG}) = .5 \)

\( \mu_{F_1,\text{POS}} = 3 \quad \mu_{F_1,\text{NEG}} = 5 \)

\[
\sigma_{F_1,\text{POS}} = \sqrt{\frac{(4-3)^2+(2-3)^2}{2}} = 1
\]

\[
\sigma_{F_1,\text{NEG}} = \sqrt{\frac{(4-5)^2+(6-5)^2}{2}} = 1
\]

\( \mu_{F_2,\text{POS}} = 4 \quad \mu_{F_2,\text{NEG}} = 2 \)

\[
\sigma_{F_2,\text{POS}} = \sqrt{\frac{(4-4)^2+(4-4)^2}{2}} = 0
\]

\[
\sigma_{F_2,\text{NEG}} = \sqrt{\frac{(2-2)^2+(2-2)^2}{2}} = 0
\]
2. Suppose you have trained three support vector machines, \(h_1, h_2,\) and \(h_3,\) returning binary classifications (+1 or −1). The observed accuracy of each of the SVMs is 90%. Assuming that the errors of these SVMs 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>((.9)^3 = .729)</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>I</td>
<td>C</td>
<td>((.9)^2(.1) = .081)</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>C</td>
<td>C</td>
<td>((.9)^2(.1) = .081)</td>
</tr>
<tr>
<td>I</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>((.9)^2(.1) = .081)</td>
</tr>
</tbody>
</table>

Total probability that \(H\) is correct is \(0.972\)
Useful formulas for Adaboost:

\[ \sum_{j=1}^{N} w_t(j) \delta(y_j \neq h_t(x_j)), \quad \text{where} \]

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

\[ \alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right) \]

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

3. 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 \( \varepsilon_1 \)?

**SOLUTION:**

\( \varepsilon_1 = .25 \)

(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:**

\[ \omega_2(1) = 0.25 \exp(-0.55(1)(1)) = 0.144 \]
\[ \omega_2(2) = 0.25 \exp(-0.55(-1)(-1)) = 0.144 \]
\[ \omega_2(3) = 0.25 \exp(-0.55(-1)(1)) = 0.433 \]
\[ \omega_2(4) = 0.25 \exp(-0.55(1)(1)) = 0.144 \]

\[ Z_1 = 0.865 \]
\[ w_2(1) = 0.144/0.865 = 0.166 \]
\[ w_2(2) = 0.144/0.865 = 0.166 \]
\[ w_2(3) = 0.433/0.865 = 0.501 \]
\[ w_2(4) = 0.144/0.865 = 0.166 \]

4. Suppose you have finished a run of Adaboost on a particular training set to produce classifiers \( h_1 \) and \( h_2 \), with \( \alpha_1 = 0.2 \) and \( \alpha_2 = -0.4 \). 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:**

\[ H(x) = \text{sgn}(0.2 h_1(x) - 0.4 h_2(x)) = \text{sgn}(-0.2 - 0.4) = -1 \]

Class is negative.
5. 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 (calculating $d^2$, where $d =$ Euclidean distance). In each iteration: (1) Assign data points to a cluster center. (2) Update the cluster centers.

**SOLUTION:**

**Iteration 1:**

\[
\begin{align*}
  d^2(x_1, m_1) &= 0^2 + 2^2 = 4 \\
  d^2(x_2, m_1) &= 0^2 + 3^2 = 9 \\
  d^2(x_3, m_1) &= 0^2 + 0^2 = 0 \\
  d^2(x_4, m_1) &= 8^2 + 1^2 = 65 \\
  d^2(x_1, m_2) &= 2^2 + 1^2 = 5 \\
  d^2(x_2, m_2) &= 2^2 + 2^2 = 8 \\
  d^2(x_3, m_2) &= 2^2 + 1^2 = 5 \\
  d^2(x_4, m_2) &= 6^2 + 0^2 = 36
\end{align*}
\]

Cluster membership:

$m_1: x_1, x_3$ \hspace{2cm} $m_2: x_2, x_4$

Updated cluster centers:

\[
\begin{align*}
  m_1: \frac{(0, 2) + (0, 0)}{2} &= (0, 1) \\
  m_2: \frac{(0, 3) + (8, 1)}{2} &= (4, 2)
\end{align*}
\]

**Iteration 2:**

\[
\begin{align*}
  d^2(x_1, m_1) &= 0^2 + 1^2 = 1 \\
  d^2(x_2, m_1) &= 0^2 + 2^2 = 4 \\
  d^2(x_3, m_1) &= 0^2 + 1^2 = 1 \\
  d^2(x_4, m_1) &= 8^2 + 0^2 = 64 \\
  d^2(x_1, m_2) &= 4^2 + 0^2 = 16 \\
  d^2(x_2, m_2) &= 4^2 + 1^2 = 17 \\
  d^2(x_3, m_2) &= 4^2 + 2^2 = 20 \\
  d^2(x_4, m_2) &= 4^2 + 1^2 = 17
\end{align*}
\]

Cluster membership:

$m_1: x_1, x_2, x_3$ \hspace{2cm} $m_2: x_4$

Updated cluster centers:

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

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