CS540: Homework Assignment #2

Matthew Klebenow
CSL: klebenow
Section 1
0 late days used

October 19, 2012

Contents

1 Question 1: Mutual Information

Table 1: Data

Temperature	**Windy**	**Humidity**	**Weather**
Cool	Yes	Low	Sunny
Mild	Yes	Low	Sunny
Hot	Yes	Medium	Rain
Cool	Yes	High	Rain
Cool	No	High	Rain
Mild	No	Low	Sunny
Cool	Yes	Medium	Rain
Mild	No	Medium	Rain
Hot	Yes	Medium	Sunny
Cool	No	High	Rain

1.1 Part a: What is the entropy of Weather?

Entropy is defined as

$$H(Y) = \sum_{i=1}^{k} -p_i \log_2 p_i \tag{1}$$

By applying equation 1 to weather with $i = 1$ corresponding to `Sunny` and $i = 2$ corresponding to `Rainy` in Table 1, we find that:

$$\begin{aligned} H(\texttt{Weather}) &= \sum_{i=1}^{2} -p_i \log_2 p_i \\ &= -\frac{4}{10} * \log_2 \frac{4}{10} - \frac{6}{10} * \log_2 \frac{6}{10} \\ &\approx (-0.4) \times (-1.322) - (0.6) \times (-0.737) \\ &\approx 0.971 \end{aligned}$$

Mutual Information Equations

$$H(Y|X = \nu) = \sum_{i=1}^{k} -\texttt{Pr}(Y = y_i|X = \nu) \log_2 \texttt{Pr}(Y = y_i|X = \nu) \tag{2}$$

$$H(Y|X) = \sum_{\nu} \texttt{Pr}(X = \nu) H(Y|X = \nu) \tag{3}$$

$$I(Y; X) = H(Y) - H(Y|X) \tag{4}$$

Equation 2 is a conditional entropy.
Equation 3 is a conditional entropy.
Equation 4 is the information gain (i.e. mutual information).

1.2 Part b: What is the mutual information between Temperature and Weather?

First we find the conditional entropy $H(\texttt{Weather}|\texttt{Temperature})$. Additionally, we will use the ordering $i = 1, 2, 3$ with respect to `Cool, Mild, Hot` and continue the previous order of $i = 1, 2$ with respect to `Sunny, Rainy`. In other words, the terms will correlate to `Cool, Mild` and `Hot` while the smaller entropy terms will be organized into probabilities akin to (`Sunny`, `Rain`).

$$\begin{aligned}
H(\texttt{Weather}) &\approx 0.971 \\
H(\texttt{Weather}|\texttt{Temperature}) &= \frac{5}{10} * H\left(\frac{1}{5}, \frac{4}{5}\right) + \frac{3}{10} * H\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{2}{10} * H\left(\frac{1}{2}, \frac{1}{2}\right) \\
&= \frac{5}{10} * \left(-\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5}\right) \\
&\ldots + \frac{3}{10} * \left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{2}{10} * 2\left(-\frac{1}{2}\log_2\frac{1}{2}\right) \\
&\approx \frac{5}{10} * (0.464 + 0.258) + \frac{3}{10} * (0.390 + 0.528) + \frac{2}{10} \\
&\approx 0.836 \\
I(\texttt{Weather}; \texttt{Temperature}) &= H(\texttt{Weather}) - H(\texttt{Weather}|\texttt{Temperature}) \\
&\approx 0.971 - 0.836 \\
&\approx 0.135
\end{aligned}$$

The mutual information between `Temperature` and `Weather` is 0.135.

1.3 Part c: What is the mutual information between Windy and Weather?

We will follow the same approach as part b. The large terms will be in the order `Windy=Yes`, `Windy=No`.

$$\begin{aligned}
H(\texttt{Weather}) &\approx 0.971 \\
H(\texttt{Weather}|\texttt{Windy}) &= \frac{6}{10} * H\left(\frac{3}{6}, \frac{3}{6}\right) + \frac{4}{10} * H\left(\frac{1}{4}, \frac{3}{4}\right) \\
&= \frac{6}{10} * 2\left(-\frac{3}{6}\log_2\frac{3}{6}\right) + \frac{4}{10} * \left(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right) \\
&\approx \frac{6}{10} + \frac{4}{10} * (0.5 + 0.311) \\
&\approx 0.9245 \\
I(\texttt{Weather}; \texttt{Windy}) &= H(\texttt{Weather}) - H(\texttt{Weather}|\texttt{Windy}) \\
&\approx 0.971 - 0.9245 \\
&\approx 0.0465
\end{aligned}$$

The mutual information between `Windy` and `Weather` is 0.465.

1.4 Part d: What is the mutual information between Humidity and Weather?

Same as previous. High order terms will be ordered `Low, Medium, High`.

$$
\begin{aligned}
H(\texttt{Weather}) &\approx 0.971 \\
H(\texttt{Weather}|\texttt{Humidity}) &= \frac{3}{10} * H\left(\frac{3}{3}, 0\right) + \frac{4}{10} * H\left(\frac{1}{4}, \frac{3}{4}\right) + \frac{3}{10} * H\left(0, \frac{3}{3}\right) \\
&= \frac{3}{10} * 0 + \frac{4}{10} * H\left(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right) + \frac{3}{10} * 0 \\
&\approx \frac{4}{10} * (0.5 + 0.311) \\
&\approx 0.3244 \\
I(\texttt{Weather}; \texttt{Humidity}) &= H(\texttt{Weather}) - H(\texttt{Weather}|\texttt{Humidity}) \\
&\approx 0.971 - 0.3244 \\
&\approx 0.6466
\end{aligned}
$$

The mutual information between `Humidity` and `Weather` is 0.6466.

2 Question 2: Pruning a Decision Tree

Given the tree provided and the following tuning set:

Table 2: Tuning Set

1	Round	Chocolate	Glazed	+
2	Round	Jelly	Glazed	-
3	Round	Chocolate	Glazed	+
4	Twist	Chocolate	Powdered	-
5	Twist	Chocolate	Sprinkled	-
6	Twist	Jelly	Powdered	+

2.1 Part a: Which nodes are candidates for pruning?

1. Root
2. Round
3. Twist
4. Round-Glazed
5. Twist-Jelly

2.2 Part b: For each node in (a), show the tuning set accuracy on the tree with that node pruned.

- Root
  1. Fail
  2. Pass
  3. Fail
  4. Pass
  5. Pass
  6. Fail

  Accuracy $= \frac{3}{6} = 0.5$.
- Round
  1. Fail
  2. Pass
  3. Fail

4. Fail
5. Fail
6. Pass

Accuracy $= \frac{2}{6} = 0.33$.

- Twist

1. Pass
2. Pass
3. Pass
4. Fail
5. Fail
6. Pass

Accuracy $= \frac{4}{6} = 0.67$.

- Round-Glazed

1. Pass
2. Fail
3. Pass
4. Fail
5. Fail
6. Pass

Accuracy $= \frac{3}{6} = 0.5$.

- Twist-Jelly

1. Pass
2. Pass
3. Pass
4. Fail
5. Fail
6. Fail

Accuracy $\frac{3}{6} = 0.5$.

2.3 Part c: Which is the first node that you would prune?

The first node that I would prune would be the one that results in the greatest accuracy. Therefore, I would prune the “Twist” node.

3 Question 3: SVM

3.1 Part a

Part A includes drawing and has been completed separately and attached to this document.

Decision boundary notes

Please note that the decision boundaries in plots i and ii both have a slope of -1 and pass through $(6, 6)$.
Also note that the decision boundary in plot iii is a horizontal line ($\texttt{slope} = 0$) that can be described as $y = 5$.

Support vector notes

Please note that the support vectors in plots i and ii are located at $(4, 4)$ and $(8, 8)$.
Also note that the support vectors in plot iii are located at $(4, 4)$ and $(4, 6)$

3.2 Part b: Which plots, if any, have the same decision boundary?

Plots i and ii have the same decision boundary.

3.3 Part c: Find the SVM solution $< \mathbf{W}, b >$ by hand for the first plot.

Assume the square is item $X_1 = (4, 4), y_1 = -1$ and the triangle is $X_2 = (8, 8), y_2 = 1$.

Hand-derivation

We first recall the definition of the decision boundary from Russell & Norvig (2009, p.745):

$$\{\mathbf{x} : \mathbf{w} \cdot \mathbf{x} + b = 0\} \tag{5}$$

We also recall that we are attempting to maximize the margin, which is defined as follows:

$$M = \frac{2}{\|\mathbf{W}\|} \tag{6}$$

Taking note of (6), we can find the value of the margin, M, geometrically. Specifically, since the square at $(4, 4)$ and the triangle at $(8, 8)$ are the support vectors and are already located along a line that is orthogonal to the decision boundary, the margin must be the distance between these two points. Mathematically:

$$
\begin{aligned}
M &= \texttt{Distance}(X_1, X_2) \\
&= \sqrt{\sum_{i=1}^{2} (X_{2,i} - X_{1,i})^2} \\
&= \sqrt{(8-4)^2 + (8-4)^2} \\
&= \sqrt{16+16} = \sqrt{32} \\
&= 5.6569
\end{aligned}
$$

We can plug this result of M back into (6) to find $\|\mathbf{W}\|$:

$$
\begin{aligned}
\|\mathbf{W}\| &= \frac{2}{M} \\
&= \frac{2}{5.6569} \\
&= 0.3536
\end{aligned}
$$

Now we know the norm of $\mathbf{W}$. We also know that it must be orthogonal to the decision boundary and must be 2-dimensional. Therefore, since the decision boundary has a slope of -1, we know that $\mathbf{W}$ must have a slope equal to $-\frac{1}{-1} = 1$. Put simply, the two entries in $\mathbf{W}$ ($\mathbf{W}[1]$ and $\mathbf{W}[2]$) must be equivalent.

$$
\begin{aligned}
\sqrt{\mathbf{W}[1]^2 + \mathbf{W}[2]^2} &= \|\mathbf{W}\| \\
\sqrt{\mathbf{W}[1]^2 + \mathbf{W}[1]^2} &= \|\mathbf{W}\| \\
\sqrt{2\mathbf{W}[1]^2} &= 0.3536 \\
\sqrt{2}\mathbf{W}[1] &= 0.3536 \\
\mathbf{W}[1] &= 0.25 \\
\mathbf{W}[2] &= 0.25
\end{aligned}
$$

Now that we have fully defined $\mathbf{W}$, we must find a b that satisfies both (5) and the given class labels from the problem statement.

$$
\begin{aligned}
\mathbf{W} \cdot \mathbf{X_1} + b &= -1 \\
\sum_{i=1}^{2} (\mathbf{W}[i] \times \mathbf{X_1}[i]) + b &= -1 \\
(0.25 \times 4 + 0.25 \times 4) + b &= -1 \\
(1+1) + b &= -1 \\
b &= -3
\end{aligned}
$$

Let's see if this value of b will satisfy the second point, X_2.

$$
\begin{aligned}
\mathbf{W} \cdot \mathbf{X_2} + b &= 1 \\
\sum_{i=1}^{2} (\mathbf{W}[i] \times \mathbf{X_2}[i]) + b &= 1 \\
(0.25 \times 8 + 0.25 \times 8) - 3 &= 1 \\
(2 + 2) - 3 &= 1 \\
1 &= 1
\end{aligned}
$$

Both equations have been satisfied. Thus, our SVM solution $< \mathbf{W}, b >$ is shown below as (7).

$$
\mathbf{W} = \begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix} \quad b = -3 \tag{7}
$$

References

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). New Jersey: Pearson.