TCSS 455 Machine Learning

Homework #1

Tianyi Li
Student ID: 1827924

1. Hypothesis space and inductive bias:

(a) Let's consider the hypothesis space H_1 consisting of all functions that take n such 3-valued input arguments and produce one 3-valued output. How many hypotheses are there in H_1? Briefly explain your answer.

- *$X = \{T, F, U\}^n, Y = \{T, F, U\}$.*
- *For a complete hypothesis space, there are 3^{3^n} possible functions in H_1.*

(b) How many hypotheses are there in H_2? Briefly explain your answer.

- *To illustrate with an example, say there are only 2 such 3-valued input arguments and produce 1 of the 3-valued output, e.g. there are inputs x_1, x_2 to yield an output Y.*
- *If x_1, $x_2, \dots, x_n$ is U, then $Y = U$: there are 5 combination of x_1, x_2 (listed in the truth table below) that could satisfy.*

  Truth table:

x_1	x_2	Y
T	T	
T	F	
T	U	U
F	T	
F	F	
F	U	U
U	T	U
U	F	U
U	U	U

- *Otherwise Y is equivalent to a conjunction of any subset of the variables $\{x_1, x_2, \dots, x_n\}$: as we only the cases below in the truth table, there is no combination could result in $Y = U$. Then $Y = T$ or $Y = F$*

  Truth table:

x_1	x_2	Y
T	T	T
T	F	F
F	T	F
F	F	F

- *Thus, to expand to the input of x_1, $x_2, \dots, x_n$, the possible resulting combination is 0.*
- *Therefore, there is $3^0 = 1$ possible functions in the H_2 hypothesis space.*

(c) Give a set of training examples such that there is no hypothesis in H_1 consistent with these examples, or explain why you cannot do that.

- *There is no such set, because it has the complete set of hypothesis spaces, such that every variation of the input and output are in there.*

(d) Give a set of training examples such that there is no hypothesis in H_2 consistent with these examples, or explain why you cannot do that.

-

(e) In which of the two cases (H_1 or H_2) is the inductive bias the highest? What are the implications of the choice of hypothesis space H_1 versus H_2 for a machine learning algorithm that tries to learn the unknown function f from training data?

- *H_2 has the higher inductive bias; it limited its hypothesis space by conjuncture rules.*
- *On the other hand, H_1 serves as the complete hypothesis space, where it has the complete ignorance (no bias at all).*

2. Decision Trees:

(a) Show the decision tree that would be learned by ID3 assuming it is given the four training examples for the Enjoy Sport? target concept shown in Table 2.1 of Chapter 2.

- *S = [3+, 1-], since there are 3 positive training examples and 1 negative training example.*
- *Entropy(S)* $= -\frac{3}{4}\ log_2\left(\frac{3}{4}\right) - \frac{1}{4}\ log_2\left(\frac{1}{4}\right) \approx 0.811$
- *For the attribute 'Sky':*

  Entropy(Sunny) $= 0$

  Entropy(Rainy) $= 0$

  Gain(S, Sky) $= Entropy(S) - \sum_{i \in Sky} \frac{|S_i|}{|S|} Entropy(S_i)$

  $= 0.811 - 0 - 0 = 0.811$
- *For the attribute 'Temperature':*

  Entropy(Warm) $= 0$

  Entropy(Cold) $= 0$

  Gain(S, Temp) $= Entropy(S) - \sum_{i \in Temp} \frac{|S_i|}{|S|} Entropy(S_i)$

  $= 0.811 - 0 - 0 = 0.811$
- *For the attribute 'Humidity':*

  Entropy(Normal) $= 0$

  Entropy(High) $= -\frac{2}{3}\ log_2\left(\frac{2}{3}\right) - \frac{1}{3}\ log_2\left(\frac{1}{3}\right) \approx 0.918$

$$Gain(S, Humidity) = Entropy(S) - \sum_{i \in Humidity} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0 - \frac{3}{4} \cdot 0.918 = 0.123$$

- *For the attribute '<u>Wind</u>':*

$$Entropy(Strong) = 0$$
$$Gain(S, Wind) = Entropy(S) - \sum_{i \in Wind} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0.811 = 0$$

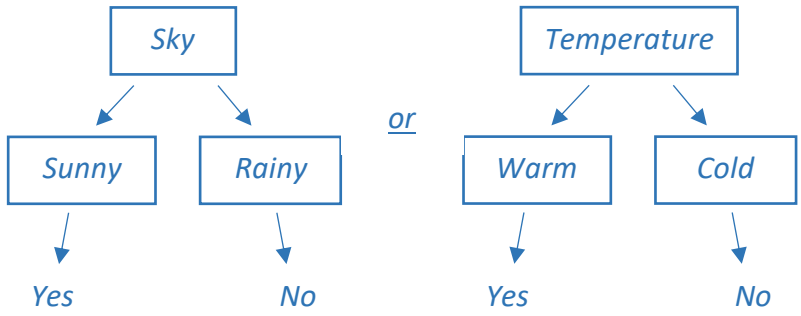
- *For the attribute '<u>Water</u>':*

$$Entropy(Warm) = -\frac{2}{3} log_2\left(\frac{2}{3}\right) - \frac{1}{3} log_2\left(\frac{1}{3}\right) \approx 0.918$$
$$Entropy(Cool) = 0$$
$$Gain(S, Water) = Entropy(S) - \sum_{i \in Water} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0 - \frac{3}{4} \cdot 0.918 = 0.123$$

- *For the attribute '<u>Forecast</u>':*

$$Entropy(Same) = 0$$
$$Entropy(Change) = -\frac{1}{2} log_2\left(\frac{1}{2}\right) - \frac{1}{2} log_2\left(\frac{1}{2}\right) = 1$$
$$Gain(S, Forecast) = Entropy(S) - \sum_{i \in Forecast} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0 - \frac{1}{2} \cdot 1 = 0.311$$

- *Because attributes "Sky" and "Temperature" both yield the highest entropy, both of them could be the resulting decision Tree:*



(b) What is the relationship between the learned decision tree and the version space that is learned from these same examples? Is the learned tree equivalent to one of the members of the version space?

- *Learned decision tree is just one of the hypotheses; the resulting one that is used to conclude all the training examples.*
- *Version space is a space that includes all the possible hypotheses.*

(c) Add the following training example, and compute the new decision tree. Show the value of the information gain for each candidate attribute at each step in growing the tree.

- *It is the same process as part (a).*
- *S = [3+, 1-] and Entropy(S)* $= -\frac{3}{5}\ log_2\left(\frac{3}{5}\right) - \frac{2}{5}\ log_2\left(\frac{2}{5}\right) \approx 0.971$
- *For the attribute 'Sky':*

  Entropy(Sunny) $= -\frac{3}{4}\ log_2\left(\frac{3}{4}\right) - \frac{1}{4}\ log_2\left(\frac{1}{4}\right) \approx 0.811$

  Entropy(Rainy) $= 0$

  Gain(S, Sky) $=$ *Entropy(S)* $- \sum_{i \in Sky} \frac{|S_i|}{|S|}$ *Entropy*(S_i)

  $= 0.971 - \frac{4}{5} \cdot 0.811 - 0 = 0.322$
- *For the attribute 'Temperature':*

  Entropy(Warm) $= -\frac{3}{4}\ log_2\left(\frac{3}{4}\right) - \frac{1}{4}\ log_2\left(\frac{1}{4}\right) \approx 0.811$

  Entropy(Cold) $= 0$

  Gain(S, Temp) $=$ *Entropy(S)* $- \sum_{i \in Temp} \frac{|S_i|}{|S|}$ *Entropy*(S_i)

  $= 0.971 - \frac{4}{5} \cdot 0.811 - 0 = 0.322$
- *For the attribute 'Humidity':*

  Entropy(Normal) $= -\frac{1}{2}\ log_2\left(\frac{1}{2}\right) - \frac{1}{2}\ log_2\left(\frac{1}{2}\right) = 1$

  Entropy(High) $= -\frac{2}{3}\ log_2\left(\frac{2}{3}\right) - \frac{1}{3}\ log_2\left(\frac{1}{3}\right) \approx 0.918$

  Gain(S, Humidity) $=$ *Entropy(S)* $- \sum_{i \in Humidity} \frac{|S_i|}{|S|}$ *Entropy*(S_i)

  $= 0.971 - \frac{2}{5} \cdot 1 - \frac{3}{5} \cdot 0.918 = 0.020$
- *For the attribute 'Wind':*

  Entropy(Strong) $= -\frac{3}{4}\ log_2\left(\frac{3}{4}\right) - \frac{1}{4}\ log_2\left(\frac{1}{4}\right) \approx 0.811$

  Entropy(Weak) $= 0$

  Gain(S, Wind)$= Entropy(S) - \sum_{i \in Wind} \frac{|S_i|}{|S|}\ Entropy(S_i)$

  $= 0.971 - \frac{4}{5} \cdot 0.811 - 0 = 0.322$
- *For the attribute 'Water':*

  Entropy(Warm) $= -\frac{1}{2}\ log_2\left(\frac{1}{2}\right) - \frac{1}{2}\ log_2\left(\frac{1}{2}\right) = 1$

  Entropy(Cool) $= 0$

  Gain(S, Water) $=$ *Entropy(S)* $- \sum_{i \in Water} \frac{|S_i|}{|S|}$ *Entropy*(S_i)

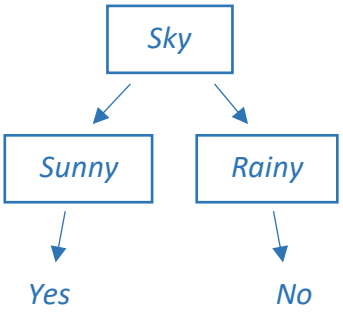
  $= 0.971 - \frac{4}{5} \cdot 1 - 0 = 0.171$
- *For the attribute 'Forecast':*

  Entropy(Same) $= -\frac{2}{3}\ log_2\left(\frac{2}{3}\right) - \frac{1}{3}\ log_2\left(\frac{1}{3}\right) \approx 0.918$

  Entropy(Change) $= -\frac{1}{2}\ log_2\left(\frac{1}{2}\right) - \frac{1}{2}\ log_2\left(\frac{1}{2}\right) = 1$

$$Gain(S, Forecast) = Entropy(S) - \sum_{i \in Forecast} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.971 - \frac{3}{5} \cdot 0.918 - \frac{2}{5} \cdot 1 = 0.002$$

- *Attributes "Sky", "Temperature" and "Wind" all have the same highest entropy, we need to choose one as the top tier of the resulting decision tree:*



- *We need to repeat the process, but eliminating the negative training example from "Sky", which means we ignore example 3 and only consider examples 1, 2, 4, 5.*
- *Then S = [3+, 1-] and Entropy(S)* $= -\frac{3}{4} log_2\left(\frac{3}{4}\right) - \frac{1}{4} log_2\left(\frac{1}{4}\right) \approx 0.811$
- *For the attribute '<u>Temperature</u>':*

$$Entropy(Warm) = -\frac{3}{4} log_2\left(\frac{3}{4}\right) - \frac{1}{4} log_2\left(\frac{1}{4}\right) \approx 0.811$$
$$Gain(S, Temp) = Entropy(S) - \sum_{i \in Temp} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - \frac{4}{4} \cdot 0.811 = 0$$

- *For the attribute '<u>Humidity</u>':*

$$Entropy(Normal) = = -\frac{1}{2} log_2\left(\frac{1}{2}\right) - \frac{1}{2} log_2\left(\frac{1}{2}\right) = 1$$
$$Entropy(High) = 0$$
$$Gain(S, Humidity) = Entropy(S) - \sum_{i \in Humidity} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - \frac{2}{4} \cdot 1 - 0 = 0.311$$

- *For the attribute '<u>Wind</u>':*

$$Entropy(Strong) = 0$$
$$Entropy(Weak) = 0$$
$$Gain(S, Wind) = Entropy(S) - \sum_{i \in Wind} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0 - 0 = 0.811$$

- *For the attribute '<u>Water</u>':*

$$Entropy(Warm) = -\frac{2}{3} log_2\left(\frac{2}{3}\right) - \frac{1}{3} log_2\left(\frac{1}{3}\right) \approx 0.918$$
$$Entropy(Cool) = 0$$
$$Gain(S, Water) = Entropy(S) - \sum_{i \in Water} \frac{|S_i|}{|S|} Entropy(S_i)$$
$$= 0.811 - 0 - \frac{3}{4} \cdot 0.918 = 0.123$$

- *For the attribute 'Forecast':*

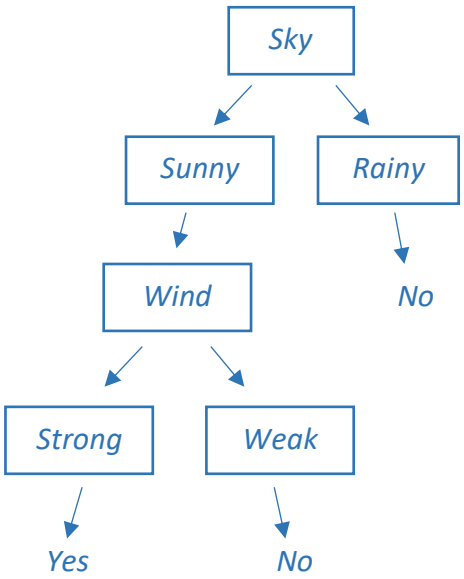
$$Entropy(Same) = -\frac{2}{3}\ log_2\left(\frac{2}{3}\right) - \frac{1}{3}\ log_2\left(\frac{1}{3}\right) \approx 0.918$$

$$Entropy(Change) = 0$$

$$Gain(S, Forecast) = Entropy(S) - \sum_{i \in Forecast} \frac{|S_i|}{|S|} Entropy(S_i)$$

$$= 0.811 - 0 - \frac{3}{4} \cdot 0.918 = 0.123$$

- *"Wind" has the highest entropy, so it would be the second tier of the decision Tree:*



- *All 5 of the training examples are concluded in the above tree, therefore we can stop and make it the final decision tree.*

(d) Suppose we wish to design a learner that (like ID3) searches a space of decision tree hypotheses and (like CANDIDATE-ELIMINATION) finds all hypotheses consistent with the data. In short, we wish to apply the CANDIDATE-ELIMINATION algorithm to searching the space of decision tree hypotheses. Show the S and G sets that result from the first training example from Table 2.1.

-

3. Machine Learning in Python

(a) Add a few lines of code to compute the accuracy in a 10-fold cross-validation set up.

```
import pandas as pd
from sklearn import tree
from sklearn import metrics
from sklearn.model_selection import cross_val_score

# Read the dataset into a dataframe and map the labels to numbers
df = pd.read_csv('iris.csv')
map_to_int = {'setosa':0, 'versicolor':1, 'virginica':2}
df["label"] = df["species"].replace(map_to_int)
#print(df)

# Separate the input features from the label
features = list(df.columns[:4])
X = df[features]
y = df["label"]

# Train a decision tree and compute its training accuracy
clf = tree.DecisionTreeClassifier(max_depth=2, criterion='entropy')
scores = cross_val_score(clf, X, y, cv=10, scoring='accuracy')
print(scores)
clf.fit(X, y)
print(metrics.accuracy_score(y,clf.predict(X)))
```

- *Added new lines are indicated above with arrows.*

(b) What is the training accuracy? What is the accuracy obtained using 10-fold cross-validation? Briefly comment on which one is the lowest, and why that does (or does not) agree with your expectations.

- *Training accuracy is 0.96.*
- *The accuracy obtained using 10-fold cross-validation is 0.96 as well.*
- *It seems like they both have the same accuracy. I expected the cross-validation accuracy would be higher than the training one. However, since the training one is high enough, such that it is close to 1.00, I would not expect cross-validation to push the accuracy to perfection.*
- *It could mean that the original training method is already close enough to yield a good model. cross-validation can be seen as a method to reassuring that.*