CSCI3230 Fundamentals of Artificial Intelligence

Written Assignment 3

Jonathan Lam – 1155093597

Question a

i.

Given a hypothesis H, its positiveness/negativeness depicts the prediction of H, whereas true/false depicts whether the nature of example agrees with the prediction of H.

	Prediction of H	In fact
TP	positive	positive
FP	positive	negative
TN	negative	negative
FN	negative	positive

ii.

Generalization happens when the current hypothesis unable to detect a **truly positive** example; thus, the extension of the hypothesis must be increased to include the examples. It is done by "loosening" the criteria by dropping (conjunctive) conditions or adding disjunctive conditions to accept the examples into the consistent space.

With original hypothesis H_1 with definition C_1 generalized to hypothesis H_2 with definition C_2,

$$\forall x(C_1(x) \Rightarrow C_2(x))$$

The opposite holds true for **specialization**, which happens when the current hypothesis unable to detect a **truly negative** example; thus, the extension of the hypothesis must be decreased to exclude the examples. It is done by "restricting" the criteria by adding (conjunctive) conditions or dropping disjunctive conditions to reject the examples out of the consistent space.

With original hypothesis H_1 with definition C_1 specialized to hypothesis H_2 with definition C_2,

$$\forall x(C_1(x) \Leftarrow C_2(x))$$

iii.

In supervised learning, a hypothesis, h, is constantly updating itself to approximates the correct value of the function, f, by the feedback of all the input-output pairs $(x, f(x))$. For h, there is many ways to approximates f; the large number of possible consistent hypotheses means 1) at any given moment, without further knowledge, there is no way to know which h is better (more consistent), and 2) there is a preference of choice *beyond* consistency with examples – such preference is called a bias.

Overfitting occurs when the representation of the function, i.e. h, adapts too well to, or too consistent with, the current problem domain (training set) and fails to predict new, unseen examples (test set) accurately. The pattern extracted by h may be correct/consistent with the examples, but it is not concise enough to encounter new examples.

Ockham's razor denotes the most likely hypothesis is the simplest one that is consistent with all observations. In other words, given two consistent hypotheses, h_1 with more conditions and h_2 with less, h_2 should adapt better when encountered with new example and is more likely to be correct then the more complex h_1. This holds true because simple hypotheses are much more common then complex ones, in which the examples may fail to meet one of the many conditions exists in h_1, causing the failure of correct prediction happening in overfitting.

Question b

i.

Entropy, or more specifically information entropy or information content, is the measure of uncertainty using probabilities. For an event V with distinct possible outcomes $v_1, \ldots, v_n$ and, for each outcome v_i, possible probability $P(v_1), \ldots, P(v_n)$, the entropy I of the actual answer is given by

$$I(V) = I(P(v_1), \ldots, P(v_n)) = \sum_{i=1}^{n} -P(v_i) log_2 P(v_i)$$

The information gain for attribute A is given by

$$\begin{aligned} Gain(A) &= I(T) - I(T|A) \\ &= I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right) \end{aligned}$$

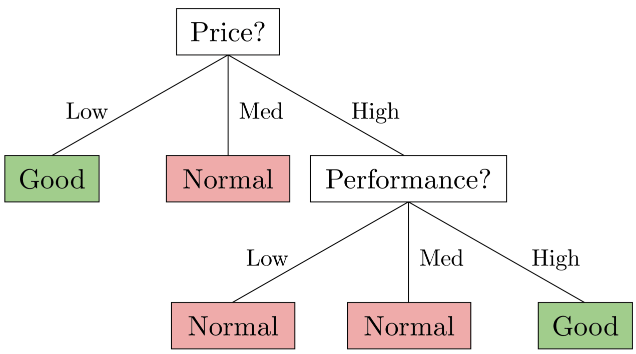
ii.

Price will be used as it divides the event into three subset, with two subsets – low and med – have definite answers of good and normal respectively. The information gain of this attribute also yields the largest compared to others.*

$$\begin{aligned} Gain(Price) &= I\left(\frac{5}{10}, \frac{5}{10}\right) - \left[\frac{5}{10} I(\frac{1}{5}, \frac{4}{5}) + \frac{1}{10} I(0, \frac{1}{1}) + \frac{4}{10} I(\frac{4}{4}, 0)\right] \\ &\approx 0.63904 \text{ bits} \end{aligned}$$

*Performance: 0.15098; Stability: 0.27549; Size: 0.02905; Service: 0.19001.

iii.



Question c

i.

R1:

$$
\begin{aligned}
Pass(x, computer) \wedge Win(x, prize) \Rightarrow Happy(x). \\
\neg(Pass(x, computer) \wedge Win(x, prize)) \vee Happy(x). \\
(\neg Pass(x, computer) \vee \neg Win(x, prize)) \vee Happy(x). \\
\neg Pass(x, computer) \vee \neg Win(x, prize) \vee Happy(x).
\end{aligned}
$$

R2:

$$
\begin{aligned}
Study(y) \vee Lucky(y) \Rightarrow Pass(y, z). \\
\neg(Study(y) \vee Lucky(y)) \vee Pass(y, z). \\
(\neg Study(y) \wedge Lucky(y)) \vee Pass(y, z). \\
(\neg Study(y) \vee Pass(y, z)) \wedge (\neg Lucky(y) \vee Pass(y, z)).
\end{aligned}
$$

R3:

$$
\begin{aligned}
Lucky(w) \Rightarrow Win(w, price) \\
\neg Lucky(w) \vee Win(w, price)
\end{aligned}
$$

Proof: $KB \models Happy(Kate)$ by resolution

1.	$\neg Pass(x, computer) \vee \neg Win(x, prize) \vee Happy(x)$	R1
2.	$\neg Study(y) \vee Pass(y, z)$	R2
3.	$\neg Lucky(y) \vee Pass(y, z)$	R2
4.	$\neg Lucky(w) \vee Win(w, price)$	R3
5.	$\neg Study(Kate)$	F1
6.	$Lucky(Kate)$	F2
7.	$\neg Happy(Kate)$	Negation of conclusion
8.	$Pass(Kate, z)$	3, 6, resolution
9.	$Win(Kate, price)$	4, 6, resolution
10.	$Happy(Kate)$	1, 8, 9, resolution
11.	$\{\}$	7, 10, resolution

ii.

Since $\forall$ is not allowed,

1. $(\neg British(Roger) \wedge Malaysian(Roger)) \vee (British(Roger) \wedge \neg Malaysian(Roger))$
2. $MainDish(Pizza) \wedge MainDish(Pasta) \wedge Golden(Pizza) \wedge Golden(Pasta)$
3. $\neg \exists x, y British(x) \wedge MainDish(y) \wedge Fruity(y) \Rightarrow Like(x, y)$
4. $\exists x MainDish(x) \wedge Golden(x) \wedge \neg Fruity(x) \Rightarrow EggFriedRice(x)$
5. $\neg \exists x Like(Roger, x) \wedge Fruity(x) \wedge EggFriedRice(x) \Rightarrow British(Roger)$