Name: CHAN King Yeung
SID: 1155119394
CSCI3230 Assignment 3

---

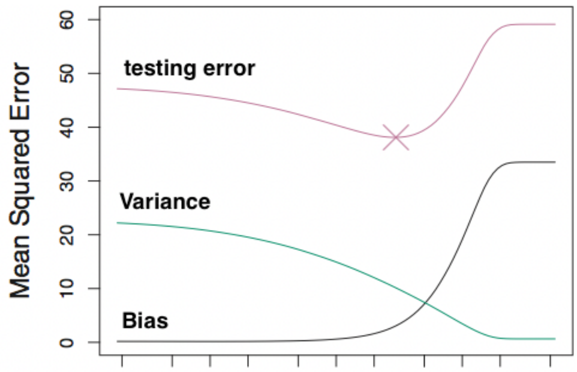
a)

i. Assume $f(x)$ is the target predicate and hypothesis H_0 predicts the predicate, say $\hat{f}(x)$. We get true positive (TP) when H_0 says $\hat{f}(x)$ is positive and $f(x)$ is positive in fact. Likewise, we get true negative (TN) when H_0 says $\hat{f}(x)$ is negative whereas $f(x)$ is negative actually. In contrast, we commit false positive (FP) if H_0 claims $\hat{f}(x)$ is positive but $f(x)$ is negative. With the same idea, we commit false negative (FN) for H_0 claims $\hat{f}(x)$ is negative but $f(x)$ is positive in fact. We can summarise the idea in the following table.

	f is positive	f is negative
$\hat{f}$ says positive	True Positive	False Positive
$\hat{f}$ says negative	False Negative	True Negative

ii. Generalisation and specialisation are methods to handle with false positive and false negative in logical formulation of learning. Let $H_0 = \{\phi_0(x) \Leftrightarrow f(x)\}$ and $H_1 = \{\phi_1(x) \Leftrightarrow f(x)\}$ for every x: $\phi_0(x) \Rightarrow \phi_1(x)$. We say H_1 is a generalisation of H_0. When H_1 commits false negative, we can drop some conditions in H_1 and it becomes H_0. Similarly, we say H_0 is a specialisation of H_1. When H_0 commits false positive, we can add some conditions in H_0 and it becomes H_1. Both generalisation and specialisation help to develop a consistent hypothesis in training data.

iii. Overfitting is a kind of error that a model over learns the structure of train data. The overfitted model only explain the idiosyncrasies in the data under study. It memorises the data pattern but not a general behaviour. In other words, it forms a lookup table instead of learning the pattern. Ockham's Razor principle is a concept which describes the simple thing is better than the complex one. When we apply the idea in inductive learning, we can interpret as simpler models usually perform better than the complicate models. We can explain the reason behind using the formula of test error which defined as $test\ error \sim bias^2 + variance + irreducible\ error$. When we develop models, the variance decreases as the degree of learning increases.



source: Lecture note from STAT4001

Simpler models usually consist more noice in learning the train data. If we apply the concept of Ockham's Razor principle, such as pruning a decision tree, we accept the model to learn more noise. From the test error formula or the above graph, we can see that increasing the variance can reduce the amount of bias which also known as bias-variance trade-off. Although simpler models have larger variance, we can reduce the test error for the model. We can utilise the principle to handle with the issue of overfitting. Just like what we have said in a second before, we can apply some technique to reduce the complexity of model, such that the variance of model will increase, and the bias will decrease. At the point where variance equals bias, the test error is minimised. If the model is underfitting, then it is another story, and we will not discuss in here.

b)

i. Entropy is a measure of impurity, which is defined as $\sum_n -\Pr(X_i)\log_2 \Pr(X_i)$. Entropy describes how much information were provided from the average information content of n events weighted by the probability of the corresponding event X_i. The greater the entropy, the little information were given from the attribute, and will lead to difficulties in prediction for unseen data. To compute the information gain for attribute A, we subtract the entropy of parent node by weight average for the entropy of child node. More formally, information gain is defined as $\mathrm{Gain}(A) = initial\ entropy - (weight\ average) \times entropy(A)$ where A is an attribute.

ii. $$\begin{aligned}\mathrm{E}(class) &= \mathrm{I}\left(\frac{1}{2},\frac{1}{2}\right)\\ &= 1\end{aligned}$$

$$\begin{aligned}\mathrm{Gain}(price) &= 1 - \left[\frac{5}{10}\mathrm{I}\left(\frac{1}{5},\frac{4}{5}\right) + \frac{1}{10}\mathrm{I}(0,1) + \frac{4}{10}\mathrm{I}(4,0)\right]\\ &= 0.639\end{aligned}$$

$$\begin{aligned}\mathrm{Gain}(performance) &= 1 - \left[\frac{2}{10}\mathrm{I}\left(\frac{1}{2},\frac{1}{2}\right) + \frac{4}{10}\mathrm{I}\left(\frac{3}{4},\frac{1}{4}\right) + \frac{4}{10}\mathrm{I}\left(\frac{1}{4},\frac{3}{4}\right)\right]\\ &= 0.151\end{aligned}$$

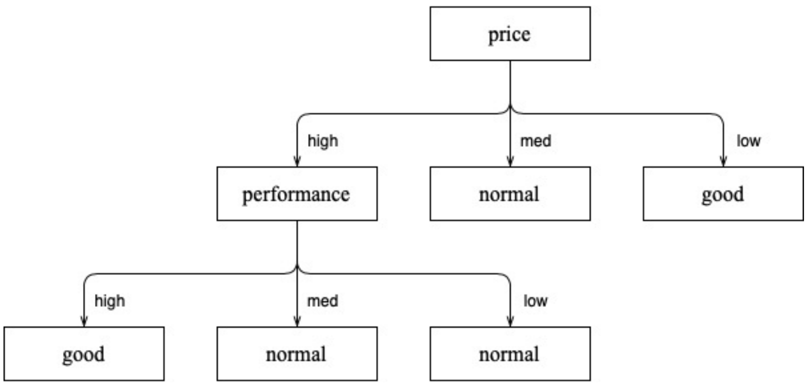
$$\begin{aligned}\mathrm{Gain}(stability) &= 1 - \left[\frac{4}{10}\mathrm{I}\left(\frac{3}{4},\frac{1}{4}\right) + \frac{4}{10}\mathrm{I}\left(\frac{1}{2},\frac{1}{2}\right) + \frac{2}{10}\mathrm{I}(0,1)\right]\\ &= 0.2755\end{aligned}$$

$$\begin{aligned}\mathrm{Gain}(size) &= 1 - \left[\frac{5}{10}\mathrm{I}\left(\frac{2}{5},\frac{3}{5}\right) + \frac{5}{10}\mathrm{I}\left(\frac{3}{5},\frac{2}{5}\right)\right]\\ &= 0.029\end{aligned}$$

$$\begin{aligned}\mathrm{Gain}(service) &= 1 - \left[\frac{4}{10}\mathrm{I}\left(\frac{1}{4},\frac{3}{4}\right) + \frac{5}{10}\mathrm{I}\left(\frac{3}{5},\frac{2}{5}\right) + \frac{1}{10}\mathrm{I}(1,0)\right]\\ &= 0.19\end{aligned}$$

Since 'price' has the largest gain, we use it to construct the root node of the decision tree.

iii.



c)

i. We convert the sentences from Horn Clause Form to Conjunctive Normal Form here.

$$\begin{aligned}
&\quad Pass(x, computer) \land Win(x, prize) \Longrightarrow Happy(x) \\
&\equiv \neg[Pass(x, computer) \land Win(x, prize)] \lor Happy(x) \\
&\equiv \neg Pass(x, computer) \lor \neg Win(x, prize) \lor Happy(x)
\end{aligned}$$

$$\begin{aligned}
&\quad Study(y) \lor Lucky(y) \Longrightarrow Pass(y, z) \\
&\equiv \neg[Study(y) \lor Lucky(y)] \lor Pass(y, z) \\
&\equiv [\neg Study(y) \land \neg Lucky(y)] \lor Pass(y, z) \\
&\equiv [\neg Study(y) \lor Pass(y, z)] \land [\neg Lucky(y) \lor Pass(y, z)]
\end{aligned}$$

$$\begin{aligned}
&\quad Lucky(w) \Longrightarrow Win(w, prize) \\
&\equiv \neg Lucky(w) \lor Win(w, prize)
\end{aligned}$$

$$\begin{aligned}
&\quad \neg Study(Kate) \\
&\equiv \neg Study(Kate)
\end{aligned}$$

$$\begin{aligned}
&\quad Lucky(Kate) \\
&\equiv Lucky(Kate)
\end{aligned}$$

Therefore, we want to show that Kate is happy i.e., $\alpha = Happy(Kate)$. To prove KB entails α, we prove by contradiction i.e., showing $KB \land \neg\alpha$ is unsatisfiable.

$$\begin{aligned}
KB \land \neg\alpha = &[Pass(Kate, computer) \land Win(Kate, prize) \Longrightarrow Happy(Kate)] \land \\
&[Study(Kate) \lor Lucky(Kate) \Longrightarrow Pass(Kate, computer)] \land \\
&[Lucky(Kate) \Longrightarrow Win(Kate, prize)] \land \\
&[\neg Study(Kate)] \land \\
&[Lucky(Kate)] \land \\
&[\neg Happy(Kate)] \\
\equiv &[\neg Pass(Kate, computer) \lor \neg Win(Kate, prize) \lor Happy(Kate)] \land \\
&\{[\neg Study(Kate) \lor Pass(Kate, computer)] \land [\neg Lucky(Kate) \lor Pass(Kate, computer)]\} \land \\
&[\neg Lucky(Kate) \lor Win(Kate, prize)] \land \\
&[\neg Study(Kate)] \land \\
&[Lucky(Kate)] \land \\
&[Happy(Kate)]
\end{aligned}$$

Consider $[\neg Pass(Kate, computer) \lor \neg Win(Kate, prize) \lor Happy(Kate)]$ and $[\neg Lucky(Kate) \lor Win(Kate, prize)]$ with complementary literals $\neg Win(Kate, prize)$ and $Win(Kate, prize)$, we have

$$\frac{[\neg Pass(Kate, computer) \lor \neg Win(Kate, prize) \lor Happy(Kate)], [\neg Lucky(Kate) \lor Win(Kate, prize)]}{\neg Pass(Kate, computer) \lor Happy(Kate) \lor \neg Lucky(Kate)}.$$

We, then, put resolvent $\neg Pass(Kate, computer) \lor Happy(Kate) \lor \neg Lucky(Kate)$ into KB.

Consider $[\neg Pass(Kate, computer) \lor Happy(Kate) \lor \neg Lucky(Kate)]$ and $[\neg Lucky(Kate) \lor Pass(Kate, computer)]$ with complementary literals $\neg Pass(Kate, computer)$ and $Pass(Kate, computer)$, we have

$$\frac{[\neg Pass(Kate, computer) \lor Happy(Kate) \lor \neg Lucky(Kate)], [\neg Lucky(Kate) \lor Pass(Kate, computer)]}{Happy(Kate) \lor \neg Lucky(Kate)}.$$

We, then, put resolvent $Happy(Kate) \lor \neg Lucky(Kate)$ into KB.

Consider $[Happy(Kate) \lor \neg Lucky(Kate)]$ and $[Lucky(Kate)]$ with complementary literals $\neg Lucky(Kate)$ and $Lucky(Kate)$, we have

$$\frac{[Happy(Kate) \lor \neg Lucky(Kate)], [Lucky(Kate)]}{Happy(Kate)}.$$

We, then, put resolvent $Happy(Kate)$ into KB.

Consider $[Happy(Kate)]$ and $[\neg Happy(Kate)]$ with complementary literals $\neg Happy(Kate)$ and $Happy(Kate)$, we have

$$\frac{[Happy(Kate)],[\neg Happy(Kate)]}{[empty]}.$$

Since the resolvent contains an empty clause, $KB \wedge \neg\alpha$ is unsatisfiable which implies $KB \Longrightarrow \alpha$ cannot be negated. Thus, KB entails α and Kate is happy.

ii.

1) $British(Roger) \vee Malaysian(Roger)$

2) $\big(MainDish(Pizza) \Longrightarrow Golden(Pizza)\big) \wedge \big(MainDish(Pasta) \Longrightarrow Golden(Pasta)\big)$

3) $\forall\, x, y\ British(x) \Longrightarrow \neg Like\Big(x, Fruity\big(MainDish(y)\big)\Big)$

HAIYAA!!!!!

4) $\exists\, x\ Golden\big(MainDish(x)\big) \wedge \neg Fruity(x) \Longrightarrow EggFriedRice(x)$

5) $\exists\, x\ Like\Big(Roger, Fruity\big(EggFriedRice(x)\big)\Big) \Longrightarrow \neg British(Roger)$