Homework 1

B99202064 Kuan Hou Chan

1 Ans: (a) (ii) (iv) (v)

(i) Programmable definition
(ii) We can find pattern of user behaviour and data from the history of card charges but the definition is not easily to program
(iii) Same with (i), which is programmable definition
(iv) Somehow there exists an underlying pattern of optimal cycle and data from a busy intersection but the definition is not easily to program
(v) Also, maybe exists an underlying pattern of the age is recommended and data from several experiment but the definition is not easily to program

2 Ans: (b) Reinforcement Learning

We use daile sales data for users and let them to rating recommendation of books which is good or bad and adjust the model to improve the performance.
Hence this is a reinforcement learning that we don't know the label but we can learn from the feedback.

3 Ans: (c) Active Learning

For automatic computer scheduling experiments, we 'asking' question strategically to mice for the effectivness of potential cancer medicines by some behaviours.
We model the schedule by mice respond to some specific questions. Hence, it is an active learning model.

4 Ans: (d) Unsupervised Learning

Without given topics, the model learning without label to categorize books into groups. Hence, this is an unsupervised learning.

5 Ans: (a) Supervised Learning

With given data that the history from each bank customer, the maximum debt comes with corresponding data features. Hence, this is an supervised learning.

6 Ans: (a) $\frac{1}{L} * (\lceil \frac{N+L}{2} \rceil - \lfloor \frac{N}{2} \rfloor)$

We can observed that $f(x)$ and $g(x)$ will result different when X_k of k is even. It is obvious that we can calculate how many even number occurs in $1 \leq k \leq N + L$,
Hence $E_{OTS}(g, f)$ is easily figure out by the number.
Since $E_{OTS}(g, f)$ is only consider the test set
$\lceil \frac{N+L}{2} \rceil$ consider upperbound of number of even number from n=0 to n=N+L, and $\lfloor \frac{N}{2} \rfloor$ consider lowerbound of number of even number from n=0 to n=N
substract these two will obtain how many even number between l=1 to l=L. For $1 \leq l \leq L$, total # of even number can be expressed as $\lceil \frac{N+L}{2} \rceil - \lfloor \frac{N}{2} \rfloor$ By $E_{OTS}(g, f)$ formula, $\Rightarrow E_{OTS}(g, f) = \frac{1}{L} * \lceil \frac{N+L}{2} - \lfloor \frac{N}{2} \rfloor \rceil$

7 Ans: (c) 2^L

For all possible $f : x \to y$: For each x, there exist 2 possible f that these 2 f's can generate D in noiseless setting if $f(x_n) = y_n$, and for the test set $\{x_l\}_l^L$, one of these two f's can generate noise y and the other can generate noiseless y. For each x_l input,we have 2 output possible and for $\{x_l\}_l^L$, we have 2*2*2*...*2 total L of 2. Hence, the total number of possible f will be 2^L.

8 Ans: (a) (c) (e)

(a) Since $E_{OTS}(g, f)$ is about test set error only, then $E_{OTS}(g, f) = \frac{k}{L}$ need k to be noise for $\sum_{l=1}^{L} g(x_{N+l} \neq f(x_{N+l}))$, for k between 0 and L, so we choose k from L and we can have the combination of LK that will satisfies the $E_{OTS}(g, f)$.
(c) $E_f\{E_{OTS}(g, f)\} = \sum_{l=1}^{L} E_{OTS}(g, f) * P(D) = \sum_{l=1}^{L} \frac{l}{L} * C_l^L(\frac{1}{2}) = \frac{1}{2}$
(e) Since all f can 'generate' D in a noiseless settings are equally likely in probablity, and by (c) that average error is about 0.5, then by the no free lunch theorem part 2: For any fixed training set D, uniformly average over f,

$$\sum_F [\varepsilon_1(E|F, D) - \varepsilon_2(E|F, D)] = 0$$

9 Ans: (a) 0.24

$\mu = 0.5$
P$(\nu = \mu) = C_5^{10} * 0.5^5 * 0.5^5 = 0.246$

10 Ans: (b) 0.39

$\mu = 0.5$
P$(\nu = \mu) = C_9^{10} * 0.9^9 * 0.1^1 = 0.387$

11 Ans: (d) $9.1 * 10^{-9}$

$\mu = 0.9$
P$(\nu <= \mu) = P(\nu = 0.1) + P(\nu = 0)$
$= C_1^{10} * 0.9^1 * 0.1^9 + C_0^{10} * 0.9^0 * 0.1^{10} = 9.1 * 10^{-9}$

12 Ans: (b) $5.52 * 10^{-6}$

$\mu = 0.9$
$\|\nu - \mu| = |0.1 - 0.9| = 0.8 > \varepsilon$
$P[0.8 > \varepsilon] \leq 2e^{-2\varepsilon^2 N} = 2e^{-2*0.8^2*10} = 5.52 * 10^{-6}$

13 Ans: (b) $\frac{8}{256}$

For each round we have 4 possible dices can choose, but for the five orange 1's condition we limited to 2 dices, B & C that match the condition.
So we only want the event that B or C occur that 2 dices to chooses.
Hence, the probability $= \frac{2*2*2*2*2}{4*4*4*4*4} = \frac{1}{32}$

14 Ans: (c) $\frac{31}{256}$

For each round we have 4 possible dices can choose, but for the 'some number' that is purely orange condition, we limited to 4 kind of combination of dices, AC AD BC BD that match the condition.
So we only want the event that the 4 combination mentioned above, then in each round we just have 2 dices to chooses and we have four kind of combination can be used then we multiply by 4.
In calculation, we recalculate the 4 combinations of AA BB CC DD then we need to substract it in the result.
Hence, the probability $= 4 * \frac{2*2*2*2*2}{4*4*4*4*4} - \frac{4}{4*4*4*4*4} = \frac{31}{256}$

15 Bonus

$R^2 = max_n \|x_n\|^2$
if X_n scale down by a factor 10, $R^{'2} = max_n \|\frac{x_n}{10}\|^2 = \frac{1}{100} * max_n \|x_n\|^2 = \frac{1}{100} * R$
Similarly for $\rho^{'2} = \frac{\rho}{100}$
Since mistake correction T will be bounded by $T \leq \frac{R^2}{\rho^2}$
then $T^{'} \leq \frac{R^{'2}}{\rho^{'2}} = \frac{R^2}{100*} * \frac{100}{\rho^2} = T$
Hence, Dr. learn plan won't work.