1. **[1+1 points]** Identify the following as a valid argument or a fallacy. If the argument is valid, identify the argument form.
Erin spends the summer in Europe but she spends the school year (fall, winter and spring) in the United States to go to school. Therefore, Erin spends the year in Europe or in the United States.

   (a) Modus Ponens
   (b) Moduls Tollens
   (c) Resolution → needs explanation.
   (d) Hypothetical syllogism
   (e) None of these

1

2. **[3 points]** If a person is wizard, it can do magic. If a person is muggle, cannot do magic. X is a muggle. Hence, X is not a wizard.
Is the above argument a valid argument?

A: ~~person~~ is a wizard — persa
B: ~~person~~ is muggle — persa
C: ~~person~~ can do magic — persa

S1: A → C
S2: B → ¬C
S3: B
S4: C → ¬B (Moduls tollens on S2)
S5: A → ¬B (Hypothetical syllogism on S1, S3)
S6: ~~B~~ B → ¬A (Modus tollens on S5)
S7: ¬A (S3, S6)

Assume person = X

∴ Above argument is valid

3

3. **[2+2 points]** Let p and q be two propositions. Consider the following two formulae in propositional logic.

$\vee$ = OR
$\wedge$ = AND

$$S1: (\neg p \wedge (p \vee q)) \rightarrow q$$
$$S2: q \rightarrow (\neg p \wedge (p \vee q))$$

Using truth tables, argue whether S1 is Tautology or not?
Same for S2?

S1 can be represented as $\neg(\neg p \wedge (p \vee q)) \vee q$

p	q	$p \vee q$	$\neg p$	$\neg p \wedge (p \vee q)$	$\neg(\neg p \wedge (p \vee q))$	S1
0	0	0	1	0	1	1
0	1	1	1	1	0	1
1	0	1	0	0	1	1
1	1	1	0	0	1	1

∴ S1 is tautology ✓ 2

S2 can be represented as $\neg q \wedge (\neg p \wedge (p \vee q))$

p	q	$\neg p$	$p \vee q$	$\neg p \wedge (p \vee q)$	$\neg q$	S2
0	0	1	0	0	1	1
0	1	1	1	1	0	1
1	0	0	1	0	1	1
1	1	0	1	0	0	1

∴ S2 is Tautology ✓

5. **[2 points]** You are building association rules from the database. You come up with

$$AR: X \rightarrow Y$$

Prove that $Lift(X \rightarrow Y) = Lift(Y \rightarrow X)$
Use empirical probabilities are definitions and Probability-based definitions.

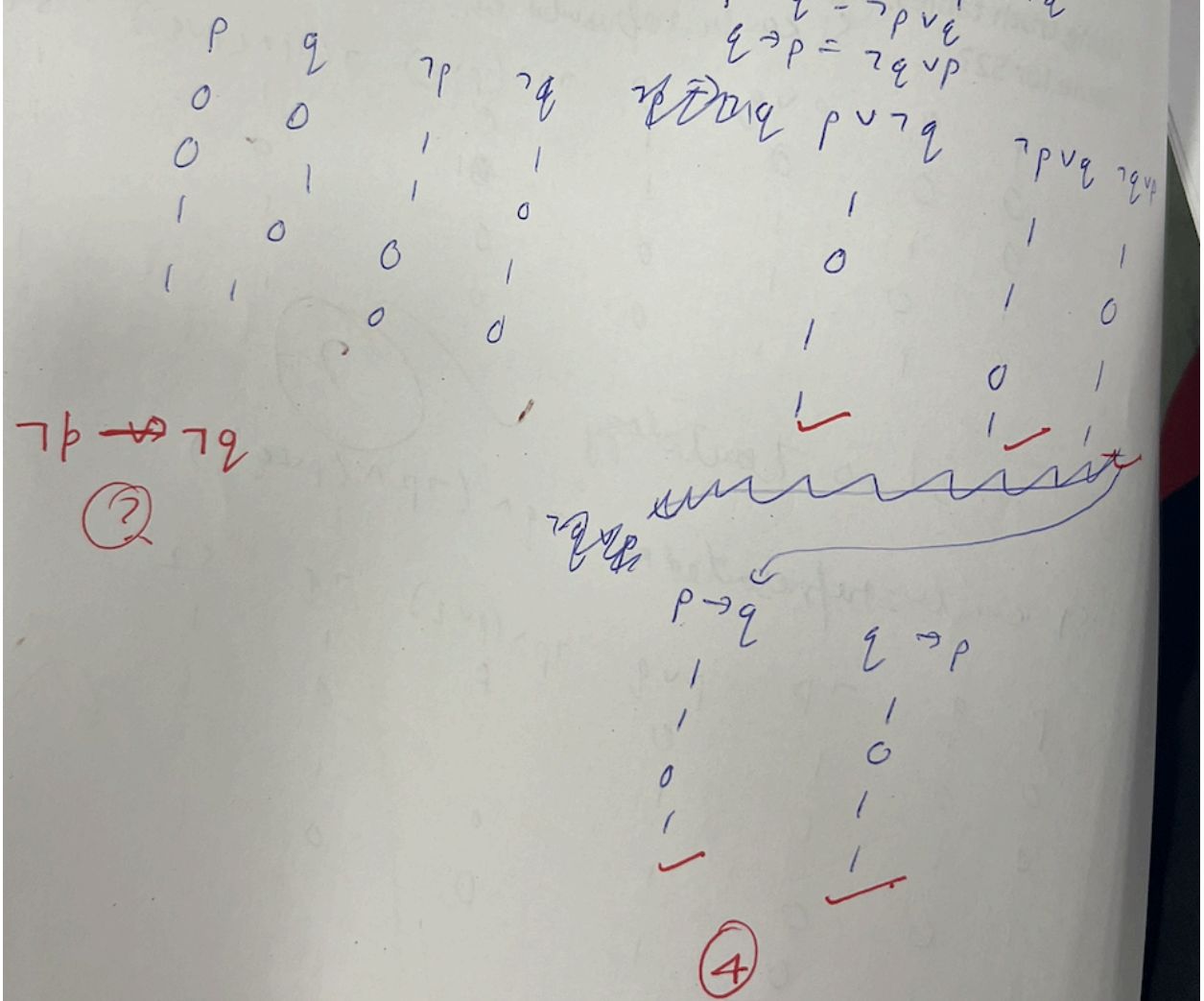
4. **[4 points]** The statement $(\neg p) \Rightarrow (\neg q)$ is logically equivalent to which of the statements below? *Use Truth Tables.*

I. $p \Rightarrow q$

II. $q \Rightarrow p$

III. $(\neg q) \vee p$

IV. $(\neg p) \vee q$



6. **[5 points]** Demonstrate how Apriori algorithm will work on the following database of transactions. Items are numbered than their actual names. Finally write down all frequent itemsets that apriori algorithm will output from singleton itemsets. Use threshold to be 2 transactions in support of the itemset to be called as frequent. ASSUME: items occurring atleast twice are frequent i.e. {2}, not {1}

Tx id	Items
1	1,2,3
2	2,3,4
3	4,5
4	1,2,4
5	1,2,3,5
6	1,2,3,4

Iteration 1: candidates {1, 2, 3, 4, 5}

~~5 gets pruned as it~~ Nothing gets pruned as all items occur atleast twice

⇒ {1, 2, 3, 4, 5} ✓ (1)

Iteration 2 candidates: {12, 23, 34, 45, 13, 14, 15, 24, 25, 35}

⇒ {12, 23, 34, 13, 14, 24}

as 45, 15, 25, and 35 get pruned ✓

Iteration 3 candidates: {123, 234, 134}

⇒ {123, 234}

as 134 gets pruned ✓ (2)

Iteration 4 candidates: ∅ as no superset exists containing all valid subsets ✓ (3)

7. Suppose you have a data tuple **DT** that has the following 5 values for temperature: **<7, 8, 6, 10, 9>.**

a. **[1+2 points]** Please compute the mean and variance of DT.

Mean : $\frac{7+8+6+10+9}{5} = \frac{40}{5} = 8$ ✓

Variance : $\frac{(8-7)^2+(8-8)^2+(8-6)^2+(8-10)^2+(8-7)^2}{5}$

$= \frac{1+0+4+4+1}{5}$

$= 2$ (3)

b. **[3+1+2 points]** Gaussian noise with a mean of 0 and variance of 2 is introduced into DT. Please create a new tuple DT' that simulates the addition of this noise. In general, we need to add large number of values for accuracy of simulation but since these are hand modeled, can add noise to reflect the needed mean and variable. Please note peak value of the noise cannot be more than 9 i.e., cannot add or subtract more than 9. Please write down the DT' vector along with the new mean and variance values. Please argue why these values are correct based on the mean and variance of the Gaussian noise added ?

~~Mean (noise, N)~~ Noise = N

Mean (N) = 0

Variance (N) = 2

DT' = { 5, 7, 6, 12, 10 }

-2 -1 0 2 1

$D_1' = \{5, 7, 6, 12, 10\}$

New mean: $\frac{5+7+6+12+10}{5} = \frac{40}{5} = 8$

New variance: $\frac{(8-5)^2 + (8-7)^2 + (8-6)^2 + (8-12)^2 + (8-10)^2}{5}$

$= \frac{9+1+4+16+4}{5} = \frac{34}{5} = 6.8$

The noise added for reference: $\{-2, -1, 0, 2, 1\}$

The new mean is correct as the mean of the noise is 0
⇒ new mean will not change

The new variance is also correct as old variance is 2 and variance of noise is also 2, with the mean as 8

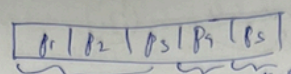
5

8. Suppose you plan to perform Nested Cross Validation where there are 5 folds of data <f1, f2, f3, f4, f5> including the Validation and Test data folds. There is a variable X which needs fine tuning between 3 values x1, x2 and x3 and a variable Y which needs fine tuning between 2 values y1 and y2. Please answer the below.

   Note: For the questions below you will need to enumerate the models that will be constructed as part of providing the clear set of steps. For example, model constructed using folds say 1,2,3 would be named as M123 while a model constructed using folds say 1,2,4,5 can be named as M1245. Now if X is set to x1 in M123 while Y is not specified, you will need to mention M123(x1)() so model is specified accurately. Please note even if you did not get the right number for the number of ML models constructed, you will receive marks for the right set of models enumerated and negative marks for enumerating wrong models.

   a. **[5 points]** How many different ML models will get constructed as part of this process ? Please provide clear set of steps for how you arrive at this number along with enumeration of all the models ?

At first, ignore the possible values of x and y. Assume you are testing for some parameter z. In nested CV, we set one fold as test and 1 as validation, with the rest being used for training:



At first, assume f_4 = testing, f_5 = validation. That is one model M_{123}. Now, we alternate the validation fold to all other possibilities within the training data ⇒ 4 models (M123, M134, M234, M124). The best of these is denoted as M_{123}. This is the inner loop (⇒ 4 models per loop). Now, the test set is shifted to another fold and this process is repeated with every fold having an opportunity of being the test set

⇒ 5 test folds * 4 models per inner loop = 20 models

Now, in place of z, considering values for x and y as we do not know if they are independent, we need to test for every combination of values: {x1y1, x2y1, x3y1, x1y2, x2y2, x3y2}

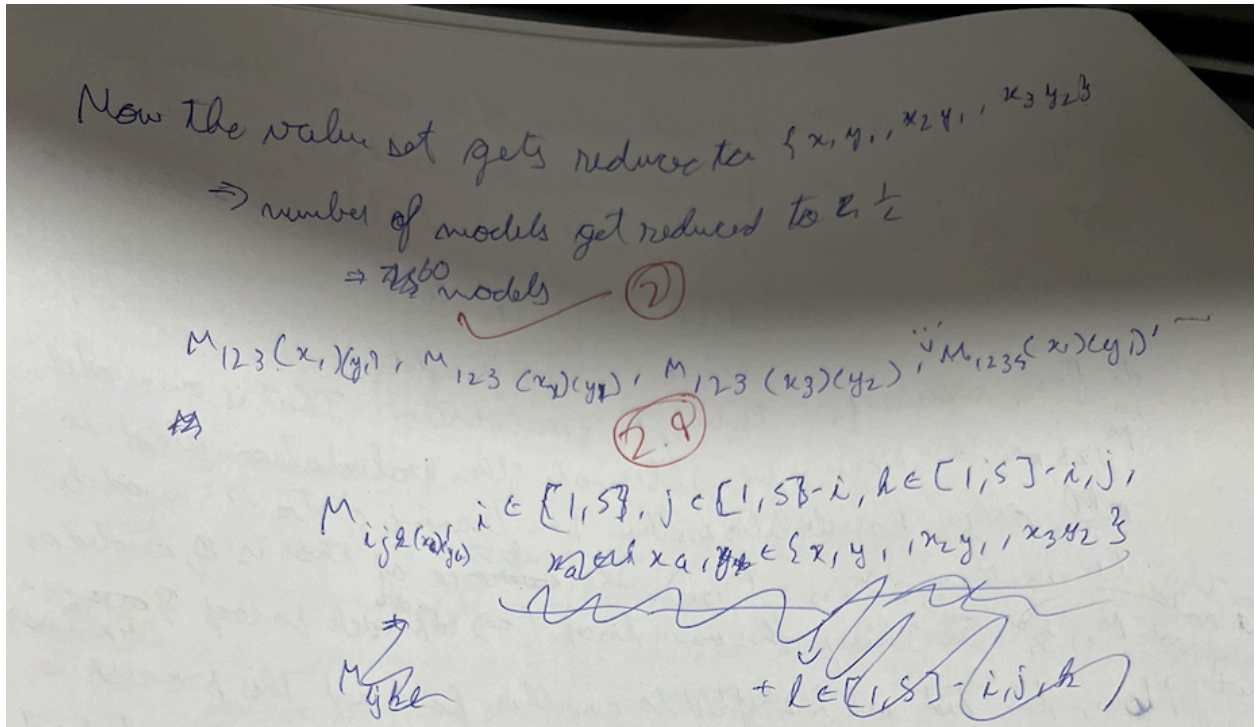
⇒ 6 values * 20 models = 120 models (2)

$M_{123(x_1)(y_1)}$, $M_{123(x_2)(y_1)}$, ..., $M_{123(x_1)(y_2)}$, ...

... $M_{234(x_1)(y_1)}$, $M_{234(x_2)(y_1)}$, ... $M_{234(x_2)(y_2)}$, ... (2.4)

M_{i3k} as defined in part (b) EXCEPT $x_a \in [x_1, x_2, x_3]$ and $y_a \in [y_1, y_2]$

b. **[5 points]** If X and Y have the following relationship wherein if Y is set to y1 one of x1 or x2 would be best for X while if y2 is selected x3 would be the best. How many ML models will get constructed ? Please provide clear set of steps for how you arrive at this number along with enumeration of all the models ?



9. You are searching for a house to buy wherein the 2 salient features that affect the price are the Location and Size of the house. Location can be Good (G) or Bad (B) while Size can be Big (B), Medium (M) or Small (S). Following are combinations of <Location, Size> mapped to Price:

<G, B>	400
<G, M>	300
<G, S>	200
<B, B>	400
<B, M>	250
<B, S>	150

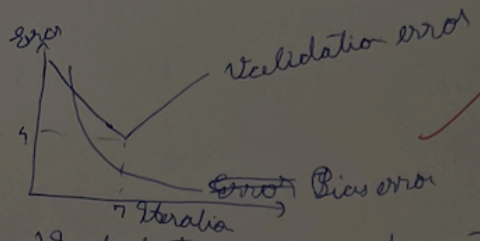
a. **[6 points]** Please compute the value of each feature along with value of constant and interaction terms if present. Please note interaction term can be used for atmost two combinations of <Location, Size>. You can present one feasible solution if multiple solutions maybe present.

b. **[4 points]** Although the above table of data is provided, you are aware that there are no Big houses in Bad locations in reality – given this information please re-compute the value of each feature along with value of constant and interaction terms if present. Please note interaction term can be used for atmost two combinations of <Location, Size>. You can present one feasible solution if multiple solutions maybe present.

c. **[3 points]** One hot encoding is used to represent the data above. Please provide one hot encoded table for the same ?

10. You are using early stopping method to perform regularization. The error for bias function is 4/X where X is number of iterations while the error from the validation set is the function |x-7|+4. Please answer the following:

a. ***[3+2 points]*** Please compute which iteration the early stopping method would recommend to stop? Why is it beneficial to stop at the iteration you computed vs. 3 steps later – please provide numerical computation to showcase the benefit obtained.



Validation error, $|x-7|+4$ is minimum when

$|x-7|=0 \Rightarrow x=7$, when it achieves a value of 4.

According to early stopping method, we should stop when validation set error is minimum

$\Rightarrow$ At $x=7^{th}$ iteration, with error $|7-7|+4=3$

For 3 iterations later, at $x=10$, error in validation set is computed as:

$|10-7|+4=7$

∴ Difference in error of validation set is $7-(|7-7|+4)$

$= 3$, meaning there is nearly twice as much error in validation set

5

b. **[3 points]** Assuming gaussian noise of mean 0 and variance σ^2 is part of dataset, would it be feasible to compute the mean square error MSE at the stopping point – please either compute the MSE or specify what additional data would be needed to compute (please also argue why it would not be feasible to derive the additional data from the provided information) ?