1. Multiple-choice questions (10 points)

Select one or more correct solutions. Please write your answer next to **Solution:**

A.) What types of learning, if any, best describe the following three scenarios:
(i) A coin classification system is created for a vending machine. In order to do this, the developers obtain exact coin specifications from the U.S. Mint and derive a statistical model of the size, weight, and denomination, which the vending machine then uses to classify its coins.
(ii) Instead of calling the U.S. Mint to obtain coin information, an algorithm is presented with a large set of labeled coins. The algorithm uses this data to infer decision boundaries which the vending machine then uses to classify its coins.
(iii) A computer develops a strategy for playing Tic-Tac-Toe by playing repeatedly and adjusting its strategy by penalizing moves that eventually lead to losing.

[a] (i) Supervised Learning, (ii) Unsupervised Learning, (iii) Reinforcement Learning
[b] (i) Supervised Learning, (ii) Not learning, (iii) Unsupervised Learning
[c] (i) Not learning, (ii) Reinforcement Learning, (iii) Supervised Learning
[d] (i) Not learning, (ii) Supervised Learning, (iii) Reinforcement Learning
[e] (i) Supervised Learning, (ii) Reinforcement Learning, (iii) Unsupervised Learning

Solution: [d] (i) Not learning, (ii) Supervised Learning, (iii) Reinforcement Learning

---

B.) For an imbalanced dataset, which of the following metric/tool is not that useful?
[a] F1 measure
[b] Accuracy
[c] Confusion Matrix
[d] Precision
Solution: [b] Accuracy

C.) Consider the following implementation of a function mysteryFunction (pseudocode), where x is a positive integer:

```
mysteryFunction (x)
    xs = str(x)
    if len(xs) == 1
        return int(xs)
    n = int(xs[0]) + int(xs[1])
    if len(xs) == 2
        return n
    else
        return n + mysteryFunction(xs[2:])
```

What does mysteryFunction(3223) return
[a] 0
[b] 10
[c] 5
[d] 1
Solution: [b] 10

D.) What is the output of the following program (in C) for input "Celestini Project"

```
#include "stdio.h"
int main()
{
   char arr[100];
   printf("%d", scanf("%s", arr));
   return 2;
}
```

[a] 0
[b] -1
[c] 1
[d] 2
Solution:[c] 1

E.) Which of the following options suggest the best approach to fix the high bias and high variance in a machine learning model? (Assume model has been trained on at least 1000 samples)
[a] To fix high bias, we can add more training samples; to fix high variance, we can reduce the number of training examples so it fits on them less
[b] To fix high bias, we can reduce our model's complexity; to fix high variance, we can increase our model's complexity
[c] To fix high bias, we can increase our model's complexity; to fix high variance, we can try reducing the number of features in the dataset
[d] To fix high bias, we can decrease the number of training samples; to fix high variance, we can increase the number of features in the dataset
Solution:[c] To fix high bias, we can increase our model's complexity; to fix high variance, we can try reducing the number of features in the dataset

---

F.) The major advantage(s) of prototyping over a Raspberry Pi over prototyping on a personal computer are
[a] cost
[b] faster processing speed
[c] small form factor
[d] low power consumption
Solution:[a],[c],[d]

G.) Which of the following statement(s) are correct?
[a] A machine learning model with higher accuracy will always indicate a better classifier.
[b] When we increase the complexity of a model, it will always decrease the test error.
[c] When we increase the complexity of a model, it will always decrease the train error.
Solution:[a] and [c]

H.) What is the output of the program (in C)?

```
#include <stdio.h>
int main()
```

```
{
int celestini[6] = {6,5,4,3,2,1};
int *ptr = (int*)(&celestini+1);
printf("%d %d", *(celestini+1), *(ptr-1));
return 0;
}
```

[a] 5 1
[b] 4 3
[c] 6 4
[d] 5 3

Solution:[a] 5 1

I.) A poor binary classification model for detecting a **rare** cancer disease *always* predicts positive for presence of the disease. What can we infer about the model's performance?

[a] The model has high accuracy, maximum precision but low recall.
[b] The model has poor accuracy, poor precision but maximum recall.
[c] The model has poor accuracy, maximum precision and minimum recall.
[d] The model has maximum accuracy, maximum precision but minimum recall.

Solution:[b] The model has poor accuracy, poor precision but maximum recall.

J.) Which of the following problems are best suited for a machine learning approach?

(i) Classifying numbers into primes and non-primes.
(ii) Detecting potential fraud in credit card charges.
(iii) Determining the time it would take a falling object to hit the ground.
(iv) Determining the optimal cycle for traffic lights in a busy intersection.

[a] (ii) and (iv)
[b] (i) and (ii)
[c] (i), (ii), and (iii).
[d] (iii)

Solution:[a] (ii) and (iv)
