Problem 1: Binary, Decimal, and Hex Number Representations [6 marks]

A conversion table is provided at the end of this exam, you can pull it off and use it for this question.

1. [1 mark] Convert the following binary number into hex: 1 1111 0000 1011
   1F0B

2. [1 mark] Convert the following hex number into decimal: 1CF
   463

3. [1 mark] Convert the following decimal number into binary: 201
   1100 1001

4. [1 mark] What is the complement of 1110 0110?
   0001 1001

5. [1 mark] What is the sum (in binary) of the following two binary numbers?
   0001 1111 + 0011 1000
   0101 0111

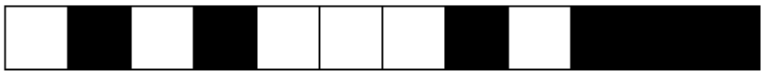
6. [1 mark] In a binary, or base-2, system, each bit, or digit, is either 0 or 1, and we can represent the eight decimal numbers from 0 to 7 using three-bit binary numbers. Consider instead a ternary, or base-3, system in which each digit is 0, 1, or 2. How many decimal numbers can be represented using three-digit ternary numbers?
   27 decimal numbers (or, the numbers from 0 to 26)

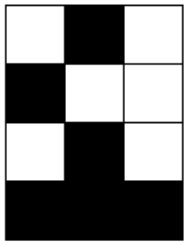
Problem 2: Image Representation [7 marks]

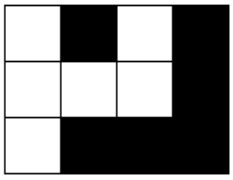
1. [1 mark] What are the three colours that make up all colours that you see on monitors and TV screens?
    a. Red, Blue, Yellow
    b. Red, Green, Blue
    c. Red, Yellow, Blue
    d. Yellow, Green, Blue

2. [1 mark] Suppose I had a colour (#456545) and then I changed it to (#006500). What happened to the colour?
    a. It stayed the same
    b. It changed into a different colour
    c. It got lighter
    d. It got darker

3. [1 mark] Given the following bitmap image representation (following the conventions described in the lectures), which is the correct image?
    3x4
    FFFFFF 000000 FFFFFF 000000 FFFFFF FFFFFF FFFFFF 000000 FFFFFF 000000 000000 000000

    a. 

    b. 

    c.  ← Answer

4. [2 marks] Give **two** situations where vector representation of data might be a better choice than bitmap representation.

   Possible answers:
   - When there's a need to keep the size of the representation small
   - When an image should look good even if you scale it up (no unwanted pixelation effects)
   - When the data is text, a vector representation ensures that text search is possible
   - When an image is composed of simple geometric features, e.g., comic book illustrations, vector representation works well

5. [1 mark] If I have a 2D image that has a large file size and becomes pixelated when I zoom in, is it more likely to be a bitmap image or vector image representation? Circle your answer.

   Bitmap Vector

6. [1 mark] In run-length-encoding, "runs" of identical colour intensities in a bitmap image representation are collapsed. Is this a lossless or lossy image compression technique? Circle your answer.

   Lossless Lossy

Problem 3: Apriori Algorithm [10 marks]

Transaction #	Items
1	Apple, banana
2	Milk, bread, orange, banana
3	Milk, chicken, bread
4	Cucumber, apple, broccoli, milk
5	Chicken, apple, milk, soda, bread
6	Milk, bread, chicken, orange, apple
7	Bread, chicken, milk, orange

1. [4 marks] What are the itemsets with 4/7 support?

Itemsets with support of 4/7 or greater:

- [0.5 marks] Apple
- [0.5 marks] Milk
- [0.5 marks] Bread
- [0.5 marks] Chicken
- [0.5 marks] {Milk, Bread}
- [0.5 marks] {Milk, Chicken}
- [0.5 marks] {Bread, Chicken}
- [0.5 mark] {Milk, Bread, Chicken}

Itemsets with support of exactly 4/7:

- [0.5 marks] Apple
- [0.5 marks] Chicken
- [1 mark] {Chicken, Milk}
- [1 mark] {Chicken, Bread}
- [1 mark] {Milk, Chicken, Bread}

2. [1 mark] Would your list of itemsets of part 1 change if transaction #7 was changed to "Orange, chicken, milk, bread, soda, banana"? Circle your answer.

Yes No

3. [1 mark] What is the confidence of Apple → {Milk, Bread}?
   2/4

4. [1 mark] What is the confidence of {Milk, Bread} → Apple?
   2/5

5. [3 marks] List all the association rules of the form "Chicken → …" that have confidence at least 2/4. (Each incorrect answer will be penalized 1 mark.)

1 mark for:

- Chicken → Milk
- Chicken → Bread
- Chicken → Apple
- Chicken → Orange

1 mark for:

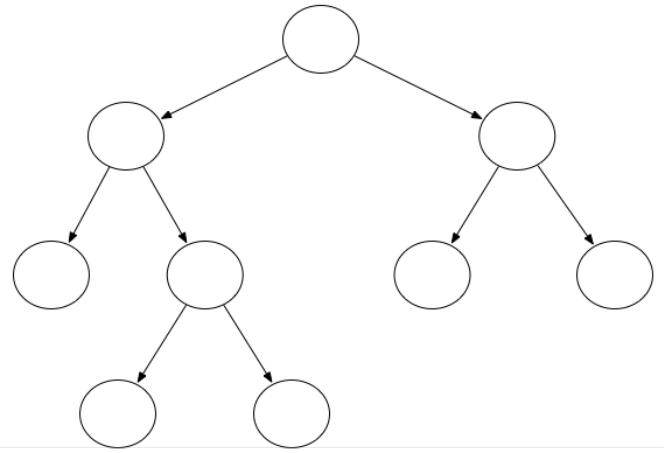
- Chicken → {Milk, Bread}
- Chicken → {Milk, Apple}
- Chicken → {Milk, Orange}
- Chicken → {Bread, Apple}
- Chicken → {Bread, Orange}

1 mark for:

- Chicken → {Milk, Bread, Apple}
- Chicken → {Milk, Bread, Orange}

Problem 4: Decision Tree [8 marks]

1. [3 marks] Draw a binary tree with a depth of 3 and with 5 leaves. You do not need to label nodes and edges.
   1 mark for a binary tree
   1 mark for the correct depth
   1 mark for the correct number of leaves



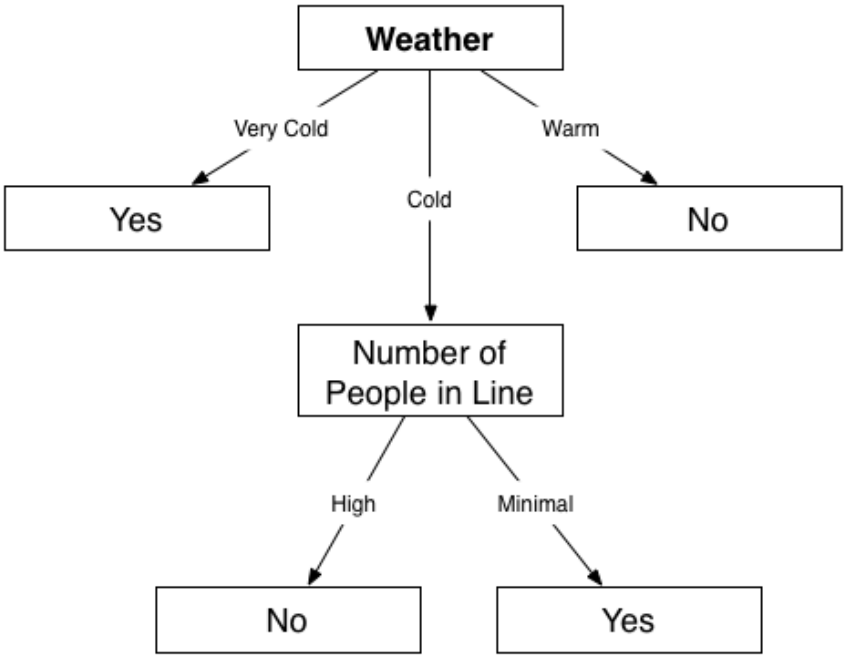
2. Abby is trying to decide whether she will go skiing today or not.

Homework Load	Number of People in Line	Weather	Ski?
Low	Minimal	Very Cold	Yes
Medium	Minimal	Warm	No
Medium	High	Very Cold	Yes
Low	Minimal	Cold	Yes
High	High	Cold	No
High	Minimal	Cold	Yes

   [3 marks] According to the definition of entropy that we used in class, what is the entropy if you split on:
   - Homework Load? 4
   - Number of People in Line? 6
   - Weather? 3

   [1 mark] What attribute should you split on first to reduce entropy?
   Weather

[1 mark] Draw the decision tree that starts splitting from the attribute you identified above.
This is marked based on the answer to the question above so even if you get the above question wrong, you are not penalized twice.



Problem 5: Clustering [5 marks]

1. [1 mark] Which of the following algorithms lists the steps of the k-means algorithm for clustering a set of points? Circle your answer.

   a. **Algorithm A**
      i. Choose k centroids at random to act as the "centre" of your clusters
      ii. For each point, determine which of the k centroids it is farthest from and assign the point to the cluster associated with that furthest centroid
      iii. Average the points inside each of the k clusters to get k new centroids
      iv. Repeat steps ii and iii as many times as you would like or until the answer stabilizes

   b. **Algorithm B**
      i. Choose k centroids at random to act as the "centre" of your clusters
      ii. Count all the points and distribute them as evenly as possible to the k centroids, to form k roughly evenly-sized clusters
      iii. Take the median of the points inside each of the k clusters to get k new centroids
      iv. Repeat steps ii and iii as many times as you would like or until the answer stabilizes

   c. **Algorithm C**
      i. Choose k centroids at random to act as the "centre" of your clusters
      ii. For each point, determine which of the k centroids it is closest to and assign this point to the cluster associated with that closest centroid
      iii. Average the points inside each of the k clusters to get k new centroids
      iv. Repeat steps ii and iii as many times as you would like or until the answer stabilizes

   d. **Algorithm D**
      i. Choose k centroids at random to act as the "centre" of your clusters
      ii. Count all the points and distribute them as evenly as possible to the k centroids, to form k roughly evenly-sized clusters
      iii. Average the points inside each of the k clusters to get k new centroids
      iv. Repeat steps ii and iii as many times as you would like or until the answer stabilizes

2. [1 mark] What is the *maximum* number of clusters that the k-means algorithm can produce, when the input is a dataset with 100 points and k = 5? Explain your answer briefly.

   5 because there is at most one cluster per centroid.

   0.5 mark for answer (5).
   0.5 mark for explanation of answer

3. [1 mark] What is the *minimum* number of clusters that the k-means algorithm can produce, when the input is a dataset with 100 points and k = 5? Explain your answer briefly.

   1 because all points could initially be closest to one of the centroids, and the resulting cluster of size 1 could be stable.

   0.5 mark for answer (1).
   0.5 mark for explanation of answer

4. [2 marks] What does it mean to say that the k-means clustering algorithm *stabilizes*?

   It means that the clusters do not change on two successive rounds of the algorithm (where a round corresponds to steps ii and iii in Algorithm C above).