**Question 1.** (9 points) There’s an old expression “red sky at night, sailor’s delight” which means that if the sky is red in the evening, then there’ll probably be good weather tomorrow. There’s also weather.gov which will provides weather forecasts by trained meteorologists. You have two hypotheses: $H_1$ = “the weather tomorrow will be good” and $H_2$ = “the weather tomorrow will be bad”. Let $D_1$ be the observation that the sky is red in the evening. Let $D_2$ be the observation that weather.gov forecasts tomorrow will be good.

1.1 (3 points): Over the last 1000 days, there have been 200 good days and 800 bad days. 160 of the good days were forecast to be good and 200 of the bad days were forecast to be good. 100 of the good days were preceded by a red sky at night and 300 of the bad days were preceded by a red sky at night. What values would you pick for the priors $P(H_1)$, $P(H_2)$ and the likelihoods $P(D_1|H_1)$, $P(D_2|H_1)$, $P(D_1|H_2)$, $P(D_2|H_2)$?

1.2 (3 points): Suppose you observe that the sky is red and the weather.gov forecasts good weather tomorrow. Using a Naive Bayes Classifier, would you deduce that the weather will be good tomorrow? Remember to show your working.

1.3 (3 points): Suppose you collect the following additional information. a) 100 of the good days that were forecast to be good were also preceded by a red sky. b) 200 of the bad days that were forecast to be good were also preceded by a red sky. Would this change your answer and why?

**Question 2.** (10 points) Consider the Markov chain with four states and transition matrix:

$$A = \begin{pmatrix}
\frac{1}{3} & \frac{2}{3} & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2}
\end{pmatrix}$$

2.1 (4 points): Draw the transition diagram. Is the Markov chain reducible or irreducible? Is the Markov chain periodic or aperiodic?

2.2 (2 points): Find the steady state distribution. Show your working.

2.3 (2 points): Suppose that the Markov chain is initially in state four. What’s the probability it is in state four after six steps? Show your working.

2.4 (2 points): Suppose that the Markov chain is observed at state one. You have two hypotheses regarding where it was at the previous step: either state one or state four. Which is the MAP hypothesis? Show your working. Hint: You should use the steady state distribution for the priors.
Question 3. (9 points) You are given a coin that may or may not be biased. Specifically, you have three hypotheses about the coin:

\[ H_1 = \text{“the coin has probability } \frac{1}{2} \text{ of landing heads”} \]
\[ H_2 = \text{“the coin has probability } \frac{1}{3} \text{ of landing heads”} \]
\[ H_3 = \text{“the coin has probability } \frac{2}{3} \text{ of landing heads”} \]

3.1 (3 points): Suppose your priors for the hypotheses are \( P(H_1) = \frac{1}{3} \), \( P(H_2) = \frac{1}{3} \), and \( P(H_3) = \frac{1}{3} \). You toss the coin seven times and observe five heads. Which is the maximum a posteriori hypothesis?

3.2 (3 points): Suppose your priors for the hypotheses are \( P(H_1) = \frac{1}{2} \), \( P(H_2) = \frac{1}{4} \), and \( P(H_3) = \frac{1}{4} \). You toss the coin seven times and observe five heads. Which is the maximum a posteriori hypothesis?

3.3 (3 points): Suppose your priors for the hypotheses are \( P(H_1) = \frac{1}{2} \), \( P(H_2) = \frac{1}{4} \), and \( P(H_3) = \frac{1}{4} \). You toss the coin seven times and observe seven heads. Which is the maximum a posteriori hypothesis?

Question 4. (11 points) Suppose that most morning you set your alarm for 8am so that you can catch the 8:30am bus to school which usually arrives just in time for your class at 9am. Unfortunately, sometimes you forget to set your alarm which might make you miss your bus and this increases the probability that you are late for class. Even if you catch the 8:30am bus, there’s a chance you’ll be late for class. To help analyze the situation we introduce some random variables. Let \( A = 1 \) if you remember to set your alarm and \( A = 0 \) otherwise. Let \( B = 1 \) if you catch the 8:30am bus and \( B = 0 \) otherwise. Let \( L = 1 \) if you are late for class and \( L = 0 \) otherwise. Suppose that \( P(A = 1) = \frac{3}{4} \), \( P(B = 1|A = 1) = \frac{4}{5} \), \( P(B = 1|A = 0) = \frac{2}{5} \), \( P(L = 1|B = 1) = \frac{1}{4} \), and \( P(L = 1|B = 0) = \frac{2}{3} \).

4.1 (1 points): What’s the values of:

\[ P(A = 0) \ , \ P(B = 0|A = 1) \ , \ P(B = 0|A = 0) \ , \ P(L = 0|B = 1) \ , \text{ and } \ P(L = 0|B = 0) \ ? \]

4.2 (1 points): Draw the Bayesian network for the random variables \( A, B, \text{ and } L \). Hint: The factorization is \( P(A = a, B = b, L = l) = P(A = a)P(B = b|A = a)P(L = l|B = b) \) for any \( a, b, l \in \{0, 1\} \).

4.3 (2 points): What’s value of \( P(A = 1, B = 1, L = 0) \)?

4.4 (2 points): What’s value of \( P(B = 1) \)?

4.5 (2 points): What’s value of \( P(L = 0) \)?

4.6 (2 points): What’s value of \( P(L = 1, A = 0) \)?

4.7 (1 points): What’s value of \( P(L = 1|A = 0) \)?
Question 5. (6 points) Suppose there are \( k \) students in a class and every student is equally likely to be born in any of the twelve months of the year. Remember to show your working for each of the following questions.

5.1 (2 points): What is the minimum value for \( k \) such that the probability there are two (or more) students born in the same month is at least \( 1/2 \).

5.2 (2 points): If \( k = 12 \), what is the probability there is a month with exactly 7 students are born?

5.3 (2 points): If \( k = 12 \), what is the probability strictly more students have birthdays in the first 6 months of the year than the second 6 months of the year.

Question 6. (8 points) In this question, the goal is to implement (in Java or Python) a 2-class Naive Bayes classifier to determine whether a city is in Russia or the US based only on the letters in the name of the city. Let \( H_1 \) be the hypothesis the city is in Russia and let \( H_2 \) be the hypothesis the city is in the US. Assume the prior probabilities are both 1/2. Let \( D_A, D_B, D_C, \ldots \) be the events that the city name contains the letter “A”, “B”, “C”, etc. You will need the following files:

https://people.cs.umass.edu/~mcgregor/240S17/russiaCities100.txt  
https://people.cs.umass.edu/~mcgregor/240S17/usCities100.txt  
https://people.cs.umass.edu/~mcgregor/240S17/russiaCitiesNext50.txt  
https://people.cs.umass.edu/~mcgregor/240S17/usCitiesNext50.txt

6.1 (3 points): Compute the 52 likelihoods \( P(D_A|H_1) \), \( P(D_B|H_1) \), \ldots \) and \( P(D_A|H_2) \), \( P(D_B|H_2) \), \ldots \). Include a printout of your code and a table of the 52 values.

6.2 (3 points): For each of the cities in \texttt{russiaCitiesNext50.txt} and \texttt{usCitiesNext50.txt}, find the MAP hypothesis. Include a printout of your code and a table containing each of the 100 cities and the MAP hypothesis. \textbf{Hint:} If the city name does not include the letter A, for example, your program will need to take into account the complement of the event \( D_A \).

6.3 (2 points): Suggest ways in which your classifier could be improved.