Bayesian Network Group Activity

Overall Learning Objectives:

- Practice data exploration techniques relevant to probabilistic modeling.
- Construct a plausible Bayesian network structure representing conditional dependencies for a given scenario.
- Identify and articulate conditional independencies implied by the network structure.
- Estimate joint and conditional probability distributions from data.
- Perform probabilistic inference (e.g., calculating the probability of an event given evidence) using the constructed network.
- Interpret and communicate the results of the Bayesian network analysis.

General Guidance for All Groups:

- Tools: You can use Python libraries like pgmpy, bnlearn, or even spreadsheet software for calculations and visualization, depending on your comfort level. Whiteboards are great for sketching initial network structures.
- Assumptions: Clearly state any assumptions you make when constructing your network (e.g., assumed causal directions not explicitly evident in the data).
- Data Limitations: Acknowledge that the provided datasets are small and simplified. Real-world applications would require larger, more nuanced data.
- Collaboration: Discuss interpretations and justifications for your network structure and findings within your group.

Group 5
Crop Failure Prediction

(Environmental Science)

- **Context:** Assess the risk of crop failure in an agricultural region based on environmental factors and farming practices.
- **Objective:** Construct a Bayesian network linking weather conditions, soil health, pest activity, farming inputs, and crop yield. Use the model to predict the probability of low yield under different scenarios and evaluate potential interventions.
- **Variables:**
    - `Rainfall` (Categorical: Low, Adequate, Excessive)
    - `Soil_Quality` (Categorical: Poor, Average, Good)
    - `Pest_Infestation` (Binary: True, False)
    - `Fertilizer_Use` (Binary: True, False)
    - `Temperature` (Categorical: Cool, Mild, Hot)
    - `Crop_Yield` (Categorical: Low, Medium, High)
- **Tasks:**

1. **Explore Data:** Examine the crop_data.csv dataset. Analyze variable distributions and look for potential relationships (e.g., how does `Rainfall` relate to `Crop_Yield`? Is `Pest_Infestation` more common with Hot `Temperature`?).
2. **Construct Network:** Design a DAG representing the causal influences. How do `Rainfall`, `Temperature`, and `Soil_Quality` affect `Crop_Yield`? How do they influence `Pest_Infestation`? Does `Fertilizer_Use` directly impact `Crop_Yield`? Justify the relationships.
3. **Estimate Probabilities:** Calculate the CPTs for each node using the provided data. For instance, determine `P(Crop_Yield=Low | Rainfall=Low, Pest_Infestation=True)`.
4. **Perform Inference:** Use the network to estimate risks and intervention effects:

    - What is the probability of Low `Crop_Yield` given Adequate Rainfall, Good `Soil_Quality`, Mild `Temperature`, no `Pest_Infestation`, and `Fertilizer_Use`?
    - How does the probability of Low `Crop_Yield` change if `Rainfall` becomes Low and a `Pest_Infestation` occurs?
    - If `Crop_Yield` is observed to be Low, what are the most probable states of Rainfall and `Pest_Infestation`?
    - Compare the probability of High `Crop_Yield` with and without `Fertilizer_Use`, assuming Average conditions otherwise.
- **Present Findings:** Summarize your network structure, key probabilities, and the outcomes of your scenario analysis.
- **Dataset:** crop_data.csv (Sample Data)