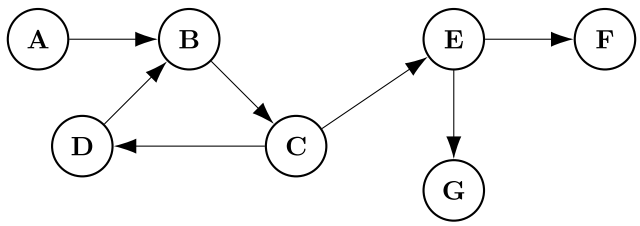
Quiz Question

Question 1: Is the graph shown below a Directed Acyclic Graph (DAG)?

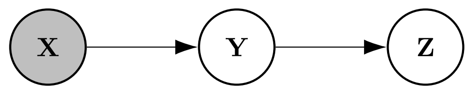


Choices:

1. Yes.
2. **No.**

Explanation: A DAG cannot contain any directed cycles. In this graph, we have the path $B \to C \to D \to B$, which is a cycle.

Question 2: Consider the causal chain shown below, where the root variable X is observed (shaded). Is the last variable Z independent of the intermediate variable Y given X? (i.e., is $Z \perp Y \mid X$?)



Choices:

1. Yes.
2. **No.**

Explanation: In a causal chain $X \to Y \to Z$, the variable Z is directly dependent on Y. Observing X influences the probability of Y, but it does not break the dependency between Y and Z. (Note: Z is independent of X only if Y is observed, which is not the case here). Also, $p(z|x, y) = p(z|y) \neq p(z|x)$.

Question 3: In a regression setting using the squared error loss $L(t, y(x)) = (t - y(x))^2$, what is the value of the expected loss (risk) when the optimal predictor $y^*(x)$ is used?
Choices:

1. **Zero.** (The loss is zero because the model is optimal).
2. **The Marginal Variance:** $\mathrm{Var}(t)$.
3. **The Expected Conditional Variance:** $\mathbb{E}_x[\mathrm{Var}(t|x)]$.
4. **The Bias squared:** $(\mathbb{E}[t] - t)^2$.

Correct Answer: 3
Explanation: The optimal predictor $y^*(x) = \mathbb{E}[t|x]$ eliminates the reducible error. However, we cannot eliminate the intrinsic noise in the data. The remaining expected loss is the variance of the target t around its mean for a given x, averaged over all x: $\mathbb{E}_x[\mathbb{E}_{t|x}[(t - \mathbb{E}[t|x])^2]] = \mathbb{E}_x[\,Var(t|x)]$.

Question 4: In a binary classification problem minimizing the misclassification rate, we assign an input vector x to the decision region $\mathcal{R}_1$ (i.e., we predict Class C_1) if specific conditions are met. Which of the following conditions correctly define this region?

Choices:

1. When the posterior probability of C_1 is maximal:
$$P(C_1|x) > P(C_2|x)$$
2. When the joint probability of x and C_1 is maximal:
$$p(x|C_1)P(C_1) > p(x|C_2)P(C_2)$$
3. When the class-conditional likelihood of C_1 is maximal:
$$p(x|C_2) > p(x|C_1)$$
4. When the prior probability of C_1 is maximal:
$$P(C_1) > P(C_2)$$

Correct Answer: 1

Explanation: To minimize misclassification, we must choose the class that is most probable given the data x. This corresponds directly to Option 1 (Posterior).

Question 5 (Medical Decision Making): A doctor is seeing a patient who presents with symptoms that could correspond to either *Disease A* or *Disease B*. The doctor has three possible actions:

- **Treat A:** Prescribe strong medication specific to Disease A.
- **Treat B:** Prescribe strong medication specific to Disease B.
- **Test:** Order a biopsy.

The loss matrix (representing harm to the patient/cost) is:

	Patient has A	**Patient has B**
Treat A	0	80
Test (Biopsy)	35	35
Treat B	50	0

If the symptoms are ambiguous and the probability of the patient having Disease A is $P(A) = 0.5$, what is the optimal decision to minimize expected harm?
Choices:

1. Treat A.
2. Test (Biopsy).
3. Treat B.
4. Treat A and Treat B are equally safe.

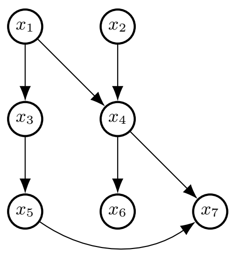
Explanation: The expected loss for A is high due to the uncertainty: $0.5(0) + 0.5(80) = 40$*. The expected loss for the Biopsy is fixed at 35. For B, we get* $0.5(50) + 0.5(0) = 25$*, which is the minimum.*

Question 6: Consider a joint distribution over 7 random variables that factorizes as follows:

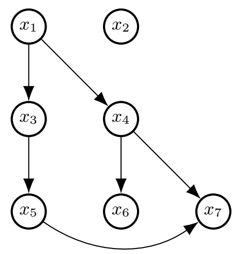
$$P(x_1, \ldots, x_7) = P(x_1)P(x_2)P(x_3|x_1)P(x_4|x_1, x_2)P(x_5|x_3)P(x_6|x_4)P(x_7|x_4, x_5)$$

Which of the following Directed Acyclic Graphs (DAGs) correctly represents this factorization?

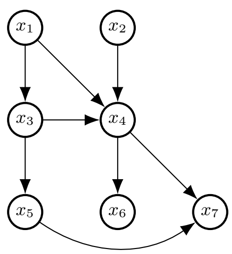
1. Graph A



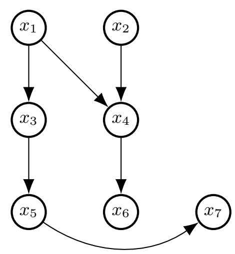
2. Graph B



3. Graph C



4. Graph D



Correct Answer: Graph A

Quiz Question 7

Question 7: Assume we have three binary random variables X, Y, Z (**taking values in** $\{0, 1\}$). We wish to parameterize their joint distribution according to one of these three different Directed Acyclic Graphs (DAGs):

1. **Causal Chain:** $X \to Y \to Z$
2. **Common Cause (Fork):** $X \leftarrow Y \to Z$
3. **Common Effect (Collider):** $X \to Y \leftarrow Z$

Which configuration requires the **most** independent parameters to define the joint distribution?
Choices:

1. The Causal Chain.
2. The Causal Chain and the Common Cause (Fork).
3. The Common Effect (Collider).
4. They all require the exact same number of parameters.

Correct Answer: 3
Explanation: Let's count the parameters for binary variables:

- **Chain (**$X \to Y \to Z$**):** $P(X)$ (1 param) + $P(Y|X)$ (2 params) + $P(Z|Y)$ (2 params) = **5 parameters**.
- **Common Cause (**$X \leftarrow Y \to Z$**):** $P(Y)$ (1 param) + $P(X|Y)$ (2 params) + $P(Z|Y)$ (2 params) = **5 parameters**.
- **Collider (**$X \to Y \leftarrow Z$**):** $P(X)$ (1 param) + $P(Z)$ (1 param) + $P(Y|X, Z)$ (4 params) = **6 parameters**.

The Collider requires 6 parameters because the child Y depends on all 4 combinations of the two parents (X, Z). The Chain and Common Cause only involve pairwise dependencies, summing to 5.