Medical Data Science
University of Tuebingen
WS 2023/2024
Atahan Özer - 6317973
Karahan Sarıtaş - 6661689

Assignment IV

Problem 1

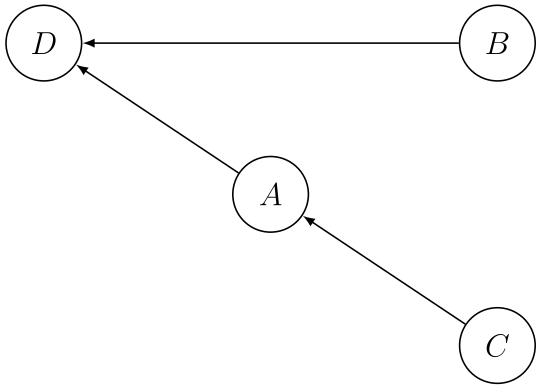
(a) Approaches that explicitly or implicitly model the distribution of inputs, as well as outputs, are known as *generative models*, because by sampling from them, it is possible to generate synthetic data points in the input space [1]. Discriminative models directly learn the probability distribution $P(Y|X)$ where Y is the class label and X is the input data. Generative models learn $P(X)$ or $P(X|Y)$ (conditional generative models) to estimate $P(Y|X)$, which allows them to generate the training data from the learned input distribution.

Bayesian networks are one of the examples of generative models that consist of *directed acyclic graphs*. In directed acyclic graphs, there is a sequence of nodes where no links extend from any node to a node with a lower number. Dependencies are taken into account in Bayes nets by drawing edges between the parent and child nodes. In the process of simulating new data from a Bayesian network, the initial step involves sampling from each of the root nodes - which we know for sure to exist. Subsequently, the sampling proceeds to the child nodes based on their respective parent(s), continuing this sequence until data for all nodes has been generated.

(b) (i) **False**. If the nodes are not d-separated, it is enough to find a single path that is not blocked by any node. If there exists a non-block path between A and B, then they are not d-separated.

(ii) **True**. We are only checking the direction of the arrows on the nodes in the set C - which is a disjoint set from B and A, by definition. Conditions to be d-separated don't impose any restrictions on the directions of edges in A and B. Therefore if A is d-separated from B by C, then B is d-separated from A by C. Another way of looking at this: A is d-separated from B by C if and only if A is conditionally independent of B given C. The latter implies that B is conditionally independent of A given C, which implies that B is d-separated from A by C.

(iii) **False**. D-separation is not transitive. We can disprove this statement with a counter-example as follows:



Here according to the definition, A is d-separated from B by C, B is d-separated from C by A, but A is not d-separated from C given any of the nodes.

(c) 1. A is directly connected to B and therefore they are **not independent**. We just have to check one path: $P_1 : A \to B$

2. We have to check the following paths:

   i. $P_1 = A \to C \to G$: The arrows on the path meet head-to-tail at C.

   ii. $P_2 = A \to B \to E \to G$: The arrows on the path meet head-to-tail at B.

   iii. $P_3 = A \to C \to E \to G$: The arrows on the path meet head-to-tail at C.

   iv. $P_4 = A \to B \to E \to C \to G$: The arrows on the path meet head-to-tail at B.
   Since all paths are blocked, we can conclude that A and G are **conditionally independent** given C, E, and B.

3. In this example, it is enough to check one path as it is not d-separated:

   i. $P_1 = D \to F \to G \to C$: The arrows meet head-to-tail at the node F but F is not in the observed set. The arrows meet head-to-head at the node G, however, G is in the observed set. Therefore this path is not blocked.
   Since there is a non-blocked path, we can say that D and C are **not conditionally independent** given E and G.

4. In this example, it is enough to check one path as it is not d-separated:

   i. $P_1 = G \to C \to E \to B$: The arrows meet head-to-tail at the node C but C is not in the observed set. The arrows meet head-to-head at the node E, however, E is in the observed set. Therefore this path is not blocked.
   Since there is a non-blocked path, we can say that G and B are **not conditionally independent** given A and E.

(d) The notation $P(A|B = b)$ represents the probability of event A occurring given that event B has occurred with a value of b. This implies that we have observed the event $B = b$. On the other hand, $P(A|\text{do}(B = b))$ signifies the probability of event A occurring when we intentionally set the variable B to a specific value, denoted as b. When we perform the intervention $\text{do}(B = b)$, we eliminate the incoming edges to the node B, indicating that $P(A|B = b)$ is not equivalent to $P(A|\text{do}(B = b))$ in the case where A is the parent node of B.

References

[1] Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag. ISBN: 0387310738.