Artificial Intelligence

Theory

Question 1

(a) **Direct Sampling**

It draws sample directly from a known distribution.

Strength ⇒ Unbiased estimate of the population
It requires no iterative refinement or additional sampling constraints. Thus it is easy and fast if the distribution is known.

Weakness ⇒ It cant handle complex relationships.

Rejection Sampling

It generates samples and rejects those that dont fit the target distribution.

Strength ⇒ It can handle constraints

Weakness ⇒ It wastes computation when many samples are rejected, especially in high dimensionality data.

Gibbs Sampling

It handles dependencies and iteratively samples each variable based on the others.

Strength ⇒ It is good for complex, interconnected data

Weakness ⇒ If the starting point is bad, it is slower and harder to set up.

Page :
Date :

(b) P(train ∩ leisure)

= P(leisure | train) × P(train)

= 0.4 × 0.3

= 0.12 .

Population sample = 100 .

Thus expected number of leisur travellers

= 0.12 × 100

= 12

(c) P(air ∩ business)

= P(air | bussiness) P(business)

= 0.2 × 0.8

= 0.16

(d) Increasing the sampling size generally increase both the accuracy and precision .

Accuracy ⇒ With more samples , the estimate tends to get closer to the true value , reducing bias .

Precision ⇒ Larger sample sizes reduce variability in the estimates , leading to more consistent results .

For the dataset :

→ On increasing the sample sizes , the estimates for probabilities like P(train ∩ leisure) or P(bus ∩ low stress) will be improved . Basically the low probability events .

Page :
Date :

→ helps minimize the uncertainity while dealing with smaller sub groups.

Question 2

(a) let $B \rightarrow$ The person reads a book.
$C \rightarrow$ Person participates in book club.
$J \rightarrow$ The person accesses the academic journals.

1. $P(J \vee B) = 0.91$
2. $P(C|B) = 0.400$ and $P(\neg J|B) = 0.6$
3. $P(C|B,J) = 0.32$ and $P(C|B,\neg J) = 0.32$
4. $P(J \wedge \neg B) = 0.227$
5. $P(\neg J \wedge \neg B) = 0.090$
6. $P(J|\neg B) = 0.716$
7. $P(C \wedge J) = 0.088$
8. $P(C \vee J) = 0.631$
9. $P(J|C) = 0.400$
10. $P(J) = 0.500$
11. $P(C|\neg B,J) = 0.0044$
$P(C|\neg B,\neg J) = 0.0044$

(b) Axiom 1 ⇒ All the probabilities are greater than 0. As they are non negative, axiom 1 is satisfied.

Axiom 2 ⇒ As $P(S) = 1$, where S is our sample space

$$P(J \cup B) + P(\neg J \wedge \neg B) = 0.91 + 0.09 = 1$$

(satisfied)

Also ~~same~~ on adding all the probabilities in the joint probability table mentioned for (c) part, we see that is sums up to 1. Therefore it also proves Axiom 2.

Page :

Date :

Axiom 3 ⟹ It states that disjoint events are mutually exclusive.
However, in our case we dont have any disjoint events.

They satisfy Axiom 1 and Axiom 2 clearly.
However we arent sure about Axiom 3 as there arent any disjoint events.

(c)

C	B	J	P.
T	T	T	0.0674
T	T	F	0.131
T	F	T	~~0.00086~~ ~~0.0006~~ 0.0009
T	F	F	0.0084
F	T	T	0.185
F	T	F	0.2788.
F	F	T	0.226
F	F	F	0.0892

(d) Checking for conditional probabilities.

(1) If J is independent ~~of~~ of B and C.

$$P(J|B,C) = P(J\cap B\cap C)\ P(B\cap C)$$
$$= P(J\cap B\cap C)(P(C\cap B\cap J) + P(B\cap J\cap C))$$
$$= (0.0674)(0.0674 + 0.131)$$
$$= 0.019 \neq P(J)$$

Thus not independent

(2) Checking conditional dependence between C and B.

$$\underset{0.2186}{P(B\cap C)} = \underbrace{\underset{0.683}{P(B)} \times \underset{0.22}{P(C)}}_{0.1502}$$

Thus $P(B\cap C) \neq P(B)P(C)$

Thus they are dependent.

Page :
Date :

(3) Similarly checking for B and J.

$P(B \cap J) = P(B)\,P(J)$

0.273 $\underbrace{0.683 \times 0.5}_{0.3415}$

$\therefore$ As $P(B \cap J) \neq P(B)\,P(J)$

They are dependent.

(4) Similarly checking for J and C.

$P(J \cap C) = P(J)\,P(C)$

0.088 $\underbrace{0.5 \times 0.22}_{0.11}$

Thus $P(J \cap C) \neq P(J)\,P(C)$

Therefore dependent.

Question 3

(a) Let, the definitions be like.

A → Adversial perturbations attack.

B → backdoor attack.

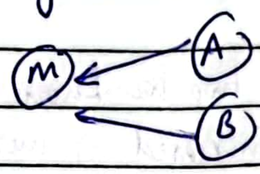
M → Misclassification alarm was trigerred.

A and B are considered to be independent of each other, ie $P(A \cap B) = P(A)\,P(B)$.

Therefore likelihood of adversial perturbations causing the misclassification is

$$P(A|M) = \frac{P(M|A)\,P(A)}{P(M)} \quad \text{(Bayes Rule)}$$

Thus, the dag looks like



(b) Prior Probabilites were $P(A)$, $P(B)$ and $P(M)$
Likelihood Probabilities are $P(M|A)$, $P(M|B)$
Posterior Probabilities are $P(A|M)$, $P(B|M)$

The priors indicate the initial probability of an event happening.
Likelihoods indicate the probability of misclassification alarm occuring given A or B occured.
While posteriors indicate the updated probability that A or B has occured given ~~and~~ an observation of misclassification alarm is taken, ~~what is the updated probability~~

$$P(A|M) = \frac{P(M|A)\,P(A)}{P(M)}, \quad P(B|M) = \frac{P(M|B)\,P(B)}{P(M)}$$

(c) $P(M) = P(M|A)\,P(A) + P(M|B)\,P(B)$

↳ Increase in $P(B)$, will ultimately increase $P(M)$.

Similarly, $P(A|M) = \frac{P(M|A)\,P(A)}{P(M)}$

here as $P(M)$ increases, rest of the probability remain same, then $P(A|M)$ decreases.

~~From the above two, we know that when backdoor triggers increased $P(B)$ will cause~~

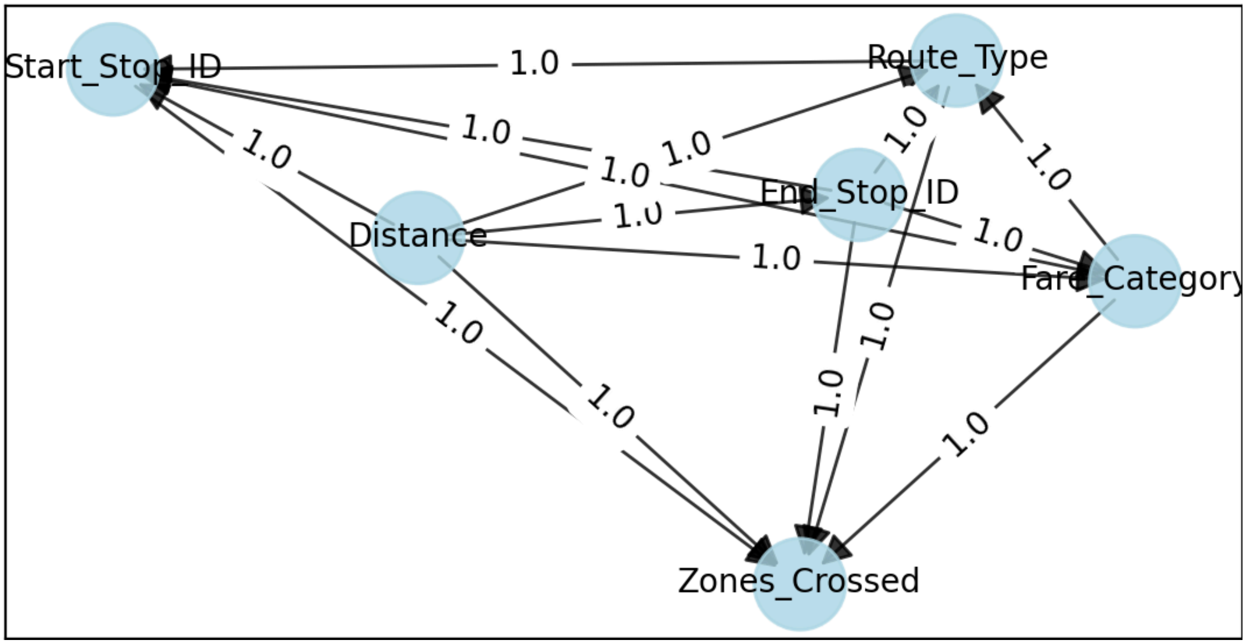
From the above two we know that when backdoor triggers are increased, missclasification alarm probability also increases.

Also if misclassification happens that is less likely due the Adversial Perturbation.

Coding Questions:

Question 4)

Question 1) c) visualization of the base model



A custom dag was created, having edges between each pair of nodes, but at the same time keeping it a acyclic graph.

Question 2)

b) The pruning method applied to the base model is the bnlearn inbuilt independence test method, which internally uses the Chi-Square test for assessing independence between variables.
The Chi-Square test evaluates the statistical independence between two variables.
Edge Pruning: Edges represent conditional dependencies between nodes. The Chi-Square test is used to evaluate the significance of each edge. If an edge fails the independence test (i.e., the dependency between the connected variables is not statistically significant), it is pruned (removed).
Node Pruning: Nodes that do not contribute meaningful information or exhibit independence from other nodes are also pruned. This reduces the overall size of the Bayesian network, streamlining computations.

The independence test helps in increasing the model's efficiency. Pruning reduces the complexity of the network by minimizing the number of nodes and edges, which decreases the time required to fit the data. A simpler network requires fewer computations during both training and inference.

```
Total Test Cases: 350
Total Correct Predictions: 350 out of 350
Model accuracy on filtered test cases: 100.00%
Time to evaluate base_model: 66.23 seconds
```

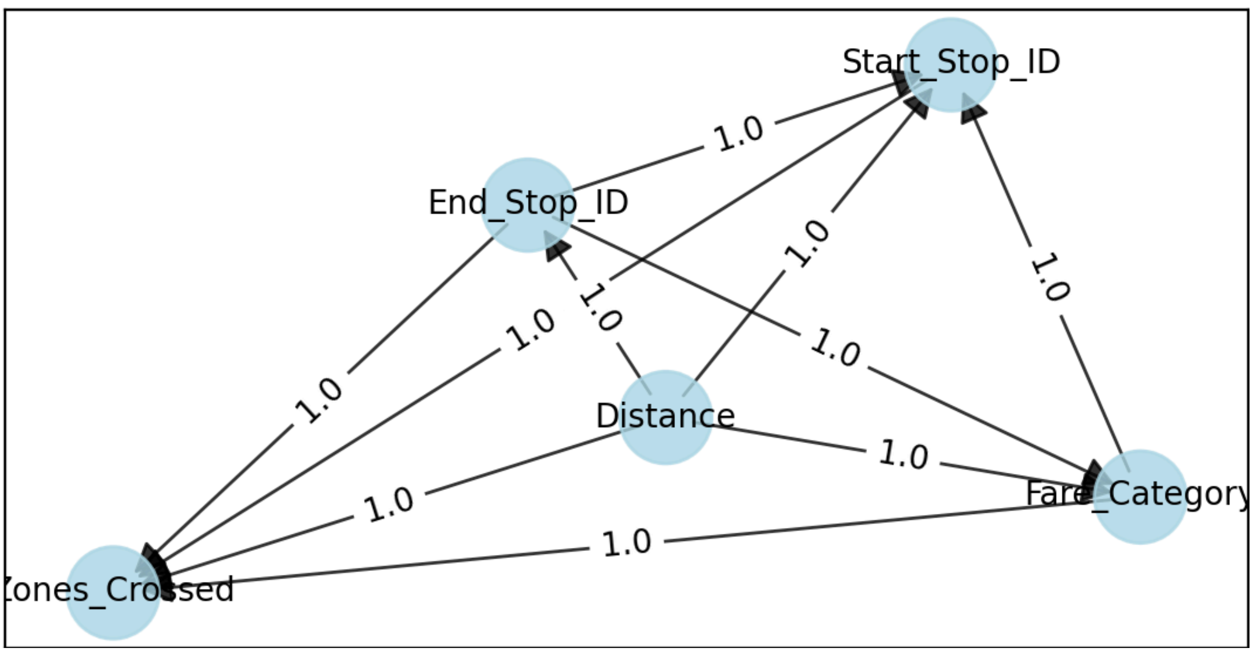
(Time required for the evaluation of the base model)

```
Total Test Cases: 350
Total Correct Predictions: 350 out of 350
Model accuracy on filtered test cases: 100.00%
Time to evaluate pruned_model: 56.92 seconds
```

(Time required for the evaluation of the pruned model)

As we can see, the time required for the evaluation of the pruned model reduced.

c)



Visualization of the pruned network. The “Route Type” node is pruned and the incoming and outgoing edges are pruned from it.

Question 3)
The optimized Bayesian network, developed using structure learning with the hill-climbing algorithm and BIC (Bayesian Information Criterion), significantly improves upon the initial network (A)
Hill Climbing: This greedy algorithm iteratively refines the base model by adding, removing, or reversing edges to improve the network structure. It ensures that the model captures meaningful dependencies directly from the data while discarding irrelevant connections.

BIC (Bayesian Information Criterion): It balances model complexity with goodness-of-fit by penalizing overfitting.

The optimized network help in increasing the accuracy as wella s efficiency as:

- Accuracy: Hill climbing extracts meaningful relationships, and BIC prevents overfitting by discarding unnecessary edges, improving predictive performance.
- Efficiency: The simplified and pruned structure reduces computational overhead, making the network faster to train and predict.

```
Total Test Cases: 350
Total Correct Predictions: 350 out of 350
Model accuracy on filtered test cases: 100.00%
Time to evaluate base_model: 66.23 seconds
```

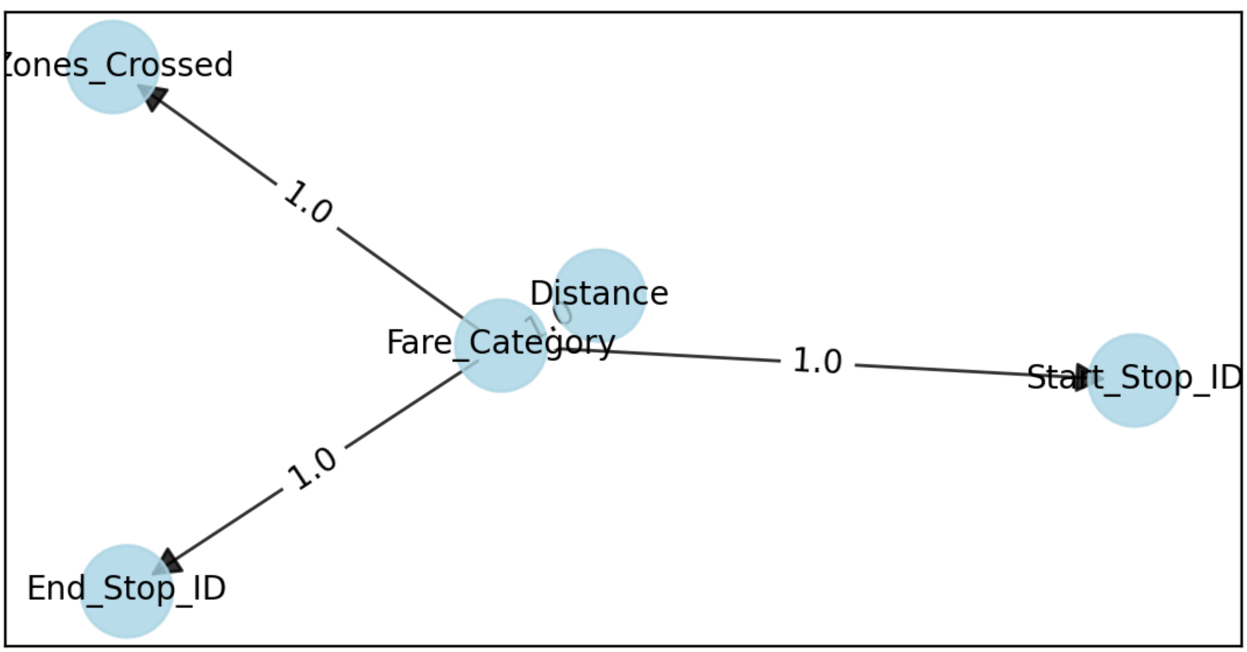
(Time required for the evaluation of the base model)

```
Total Test Cases: 350
Total Correct Predictions: 350 out of 350
Model accuracy on filtered test cases: 100.00%
Time to evaluate optimized_model: 1.32 seconds
```

(Time required for the evaluation of the optimized model)

The time required for the evaluation of the model reduced by a lot on applying structure learning.

c)



Question 5)

The **straight-until-obstacle policy** is better than the random walk policy because it relies on deterministic actions, leading to more predictable and structured movement. This reduces randomness in state transitions, making it easier to model and estimate the next state. With fewer unpredictable actions, the policy minimizes the likelihood of errors and ensures a more reliable trajectory. Additionally, its structured approach simplifies calculations, as the number of possible transitions at each step is limited. This makes the policy especially effective in handling noisy observations, as the predictable movements help counteract uncertainty, ultimately resulting in higher accuracy.