**CSEP 546: Data Mining (Spring 2017)  
Assignment 2: Collaborative Filtering and Bayesian Networks**

**-Nithin Mahesh**

**Problem 2: Bayesian Networks**

***2.1*** *You are a car mechanic and want to provide quick assesment of customers' car issues. Their car can either have a problem with the tires (low pressure, worn out, etc.) or problems with the suspension. For this problem we assume there is exactly ONE problem with the car. The customers normally report two symptoms: Is the car skidding or not and is car unstable at turns or not. From your experience you know that if the car has tire problems it is likely to skid with 50% chance while if it has suspension problems it will skid 20% of the time. You also know that a car is unstable with 30% chance if the tires are bad but with 70% if the suspension is bad. From prior experience you know that 70% of time cars have issues with their suspension and 30% with their tires.*

*a. You do a quick test drive and discover the car is unstable at turns but is not skidding. How likely is it to be tire or suspension problem? (Hint: To solve this problem, start by drawing a Bayes' net with variables C(Car issue), S(skidding) and T(stable or not at turns) and write down the conditional probability tables.)*

**Solution:**

Tabulating the given data:

|  |  |
| --- | --- |
| **Car Issue** | **Probability** |
| Tires | 0.3 |
| Suspension | 0.7 |

|  |  |  |
| --- | --- | --- |
| **Car Issue** | **Skid** | **Probability** |
| Tires | Yes | 0.5 |
| Tires | No | 0.5 |
| Suspension | Yes | 0.2 |
| Suspension | No | 0.8 |

|  |  |  |
| --- | --- | --- |
| **Car Issue** | **Stable** | **Probability** |
| Tires | Yes | 0.7 |
| Tires | No | 0.3 |
| Suspension | Yes | 0.3 |
| Suspension | No | 0.7 |

Now, let’s compute the probability of all the variables occurring together. Since Skidding is independent of Stability, it is just a product of the rows in the three tables.

|  |  |  |  |
| --- | --- | --- | --- |
| **Car Issue** | **Skid** | **Stable** | **Probability** |
| Tires | Yes | Yes | 0.105 |
| Tires | Yes | No | 0.045 |
| Tires | No | Yes | 0.105 |
| *Tires* | *No* | *No* | *0.045* |
| Suspension | Yes | Yes | 0.042 |
| Suspension | Yes | No | 0.098 |
| Suspension | No | Yes | 0.168 |
| *Suspension* | *No* | *No* | *0.392* |

Now, it is given that the car is not stable at turns but not skidding. So only the highlighted rows are of importance.

**Probability of Tire Issue: = 0.045 / (0.045 + 0.392) = 0.102974828**

**Probability of Suspension Issue: = 0.392 / (0.045 + 0.392) = 0.897025172**

*From experience you know that whenever a customer reports the problems with their car they are only 75% correct, and 25% of the time all the reported symptoms are the exact opposite of the way the car actually behaves.*

*b. What are the probabilities of the car condition given that the customer now claims that his car is skidding but otherwise stable?*

**Solution:**

Here we are adding a fourth variable whether the customer is right or wrong. For right, all rows get multiplied by 0.75 and for wrong by 0.25.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Customer** | **Car Issue** | **Skid** | **Stable** | **Probability** |
| *Right* | *Tires* | *Yes* | *Yes* | *0.07875* |
| Right | Tires | Yes | No | 0.03375 |
| Right | Tires | No | Yes | 0.07875 |
| Right | Tires | No | No | 0.03375 |
| *Right* | *Suspension* | *Yes* | *Yes* | *0.0315* |
| Right | Suspension | Yes | No | 0.0735 |
| Right | Suspension | No | Yes | 0.126 |
| Right | Suspension | No | No | 0.294 |
| Wrong | Tires | Yes | Yes | 0.02625 |
| Wrong | Tires | Yes | No | 0.01125 |
| Wrong | Tires | No | Yes | 0.02625 |
| *Wrong* | *Tires* | *No* | *No* | *0.**01125* |
| Wrong | Suspension | Yes | Yes | 0.0105 |
| Wrong | Suspension | Yes | No | 0.0245 |
| Wrong | Suspension | No | Yes | 0.042 |
| *Wrong* | *Suspension* | *No* | *No* | *0.098* |

Now the customer claims car is skidding but stable. He can either be completely right or completely wrong. So only the highlighted rows are of importance.

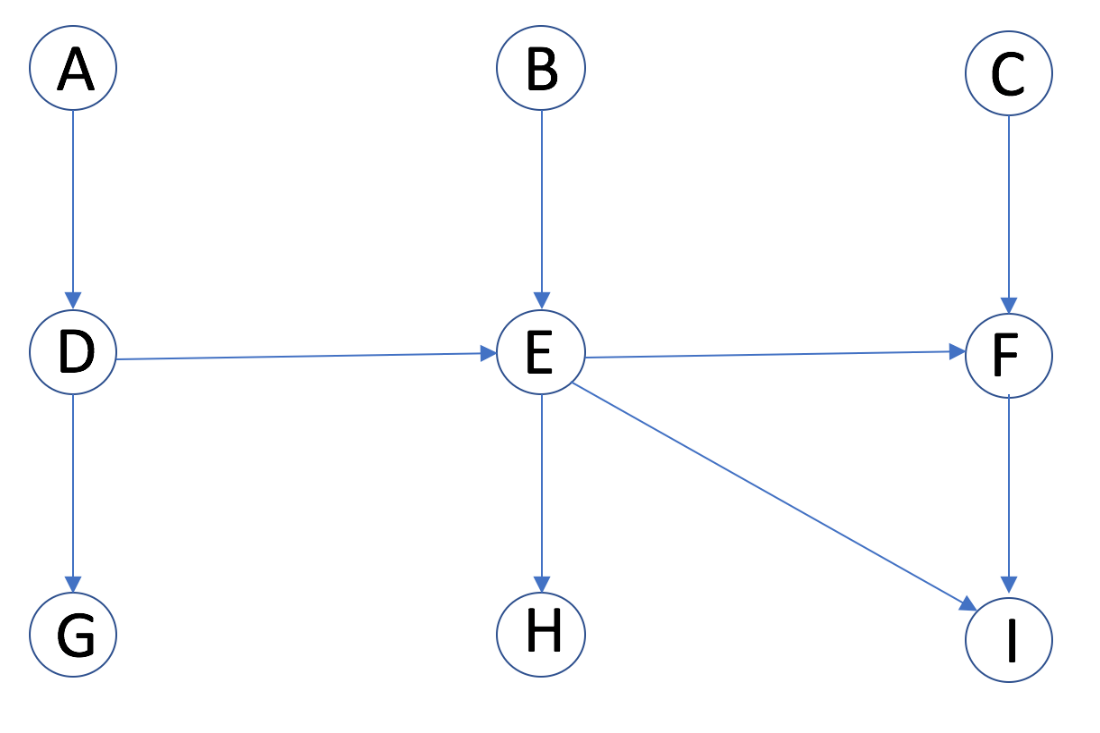
**Probability of Tire Issue:**

**= (0.07875 + 0.01125) / (0.07875 + 0.01125 + 0.0315 + 0.098) = 0.41**

**Probability of Suspension Issue:**

**= (0.0315 + 0.098) / (0.07875 + 0.01125 + 0.0315 + 0.098) = 0.59**

***2.2*** *Consider the following Bayesian network:*

**

*Which of the followings are guaranteed to be true without making any additional conditional independence assumptions, other than those implied by the graph? (Mark all true statements)*

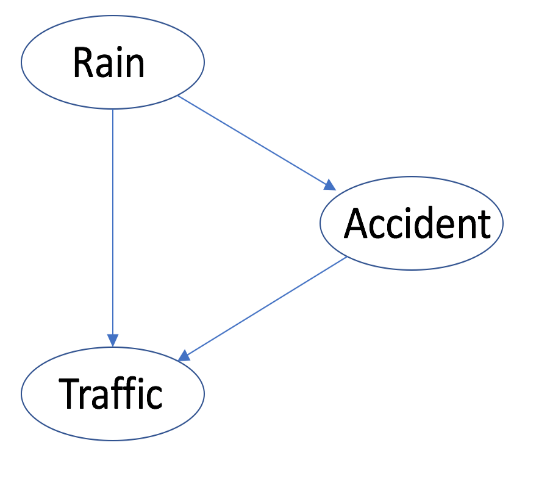
1. *P(F | E, D, A) = P(F | E)*
2. *P(G, H | A, C) = P(H | A, C) ∗ P(G | A, C)*
3. *P(B | A = a, C = c) = P(B)*
4. *P(H, I | E, F) = P(H | E, F) ∗ P(I | E, F)*
5. *Is F independent of H given B and G?*

*Justify your answers. This Bayesian network is based on Hidden Markov Model which is important tool for studying processes that evolve with time.*

**Solution:**

1. P(F | E, D, A) = P(F | E) – **TRUE**. From the network, when E is already given, probability if F does not change whatever be the values of D and A. Therefore this is true.
2. P(G, H | A, C) = P(H | A, C) ∗ P(G | A, C) – **FALSE.**  Unless, the parents of G and H are given, this is not true. Here we cannot claim G and H as independent yet. So this does not hold true.
3. P(B | A = a, C = c) = P(B) - **TRUE.** B is independent of A and C according to the network. Since B has no parent, this does not require occurrence of any prior variable.
4. P(H, I | E, F) = P(H | E, F) ∗ P(I | E, F) – **TRUE.** This is by definition from the Bayesian network principle that I is independent of H given its parents E and F.
5. Is F independent of H given B and G? – **NO.** To be claimed independent, its parents need to be given.

***2.3*** *Consider the following Bayesian network:*

**

*Any of the “Rain (R)”, “Accident (A)”, “Traffic (T)” are Boolean nodes. Suppose you have a training set composed of the following examples in the form (R,A,T), with "?" indicating a missing value: (1,1,0), (1,1,1), (1,1,1), (0,1,1), (0,1,?), (0,0,0), (0,0,?). Show the first iteration of EM algorithm (initial parameters, E-step, M-step), assuming the parameters are initialized ignoring missing values.*

**Solution:**

Following are the conditional probability tables based on given data:

|  |  |
| --- | --- |
| **Rain** | **Probability** |
| Yes | 0.43 |
| No | 0.57 |

|  |  |  |
| --- | --- | --- |
| **Accident** | **Rain** | **Probability** |
| Yes | Yes | 0.428571429 |
| Yes | No | 0.285714286 |
| No | Yes | 0 |
| No | No | 0.285714286 |

|  |  |  |
| --- | --- | --- |
| **Rain** | **Traffic** | **Probability** |
| Yes | Yes | 0.4 |
| Yes | No | 0.2 |
| No | Yes | 0.2 |
| No | No | 0.2 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Traffic** | **Rain** | **Accident** | **Probability** |
| Yes | Yes | Yes | 0.4 |
| Yes | Yes | No | 0 |
| Yes | No | Yes | 0.2 |
| Yes | No | No | 0 |
| No | Yes | Yes | 0.2 |
| No | Yes | No | 0 |
| No | No | Yes | 0 |
| No | No | No | 0.2 |

P(T/A=yes, R=no) = 0.2 / 0.2 = 1

P(T/A=no, R = no) = 0/ 0.2 = 0