AI Academy: Introduction to Data Mining
Week 7 Workshop

Workshop 7 contains 2 questions.

---

1 BN Inference (12 points)

The following dataset presents 3 categorical attributes: Gender (M, F), Car Type (Sports, Luxury) and Age Group (G1, G2) with one Class Variable: Class (C0, C1). For each question, please show how you arrived at your answer.

Gender	**Car Type**	**Age Group**	**Class**
M	Luxury	G2	C0
M	Sports	G1	C0
M	Sports	G1	C1
M	Luxury	G1	C1
M	Luxury	G2	C0
F	Sports	G1	C1
F	Luxury	G2	C1
F	Luxury	G1	C0
F	Sports	G1	C0
F	Luxury	G1	C1

Table 1: Dataset for BN Inference

For the following problem, you may find it useful to fill in the following table (optional).

$P(Class = C0) =$	$P(Class = C1) =$
$P(Gender = M \mid Class = C0) =$	$P(Gender = M \mid Class = C1) =$
$P(Gender = F \mid Class = C0) =$	$P(Gender = F \mid Class = C1) =$
$P(CarType = Luxury \mid Class = C0) =$	$P(CarType = Luxury \mid Class = C1) =$
$P(CarType = Sports \mid Class = C0) =$	$P(CarType = Sports \mid Class = C1) =$
$P(AgeGroup = G1 \mid Class = C0) =$	$P(AgeGroup = G1 \mid Class = C1) =$
$P(AgeGroup = G2 \mid Class = C0) =$	$P(AgeGroup = G2 \mid Class = C1) =$

Using the training dataset above, how would a Naive Bayes classifier classify the following data points? Show your work.

1. $\{Gender = M,\ Car\ Type = Luxury,\ Age\ Group = G1\}$
2. $\{Gender = M,\ Car\ Type = Sports,\ Age\ Group = G2\}$
3. $\{Gender = F,\ Car\ Type = Sports,\ Age\ Group = G1\}$
4. $\{Gender = F,\ Car\ Type = Luxury,\ Age\ Group = G2\}$

1. $PClass = C0|Gender = M, CarType = Luxury, AgeGroup = G1 =$
   $P(Gender = M|Class = C0) * P(CarType = Luxury|Class = C0) * P(AgeGroup = G1|Class = C0) * P(Class = C0)/P = 0.6 * 0.6 * 0.6 * 0.5/P = 0.108/P$

$P(Class = C0) = 0.5$	$P(Class = C1) = 0.5$
$P(Gender = M \mid Class = C0) = 0.6$	$P(Gender = M \mid Class = C1) = 0.4$
$P(Gender = F \mid Class = C0) = 0.4$	$P(Gender = F \mid Class = C1) = 0.6$
$P(CarType = Luxury \mid Class = C0) = 0.6$	$P(CarType = Luxury \mid Class = C1) = 0.6$
$P(CarType = Sports \mid Class = C0) = 0.4$	$P(CarType = Sports \mid Class = C1) = 0.4$
$P(AgeGroup = G1 \mid Class = C0) = 0.6$	$P(AgeGroup = G1 \mid Class = C1) = 0.8$
$P(AgeGroup = G2 \mid Class = C0) = 0.4$	$P(AgeGroup = G2 \mid Class = C1) = 0.2$

$PClass = C1|Gender = M, CarType = Luxury, AgeGroup = G1 =$
$P(Gender = M|Class = C1) * P(CarType = Luxury|Class = C1) * P(AgeGroup =$
$G1|Class = C1) * P(Class = C1)/P = 0.4 * 0.6 * 0.8 * 0.5/P = 0.096/P$

So the result is Class = C0.

2. $PClass = C0|Gender = M, CarType = Sports, AgeGroup = G2 =$
$P(Gender = M|Class = C0) * P(CarType = Sports|Class = C0) * P(AgeGroup =$
$G2|Class = C0) * P(Class = C0)/P = 0.6 * 0.4 * 0.4 * 0.5/P = 0.048/P$

   $PClass = C1|Gender = M, CarType = Sports, AgeGroup = G2 =$
$P(Gender = M|Class = C1) * P(CarType = Sports|Class = C1) * P(AgeGroup =$
$G2|Class = C1) * P(Class = C1)/P = 0.4 * 0.4 * 0.2 * 0.5/P = 0.016/P$

   So the result is Class = C0.

3. $PClass = C0|Gender = F, CarType = Sports, AgeGroup = G1 =$
$P(Gender = F|Class = C0) * P(CarType = Sports|Class = C0) * P(AgeGroup =$
$G1|Class = C0) * P(Class = C0)/P = 0.4 * 0.4 * 0.6 * 0.5/P = 0.048/P$

   $PClass = C1|Gender = F, CarType = Sports, AgeGroup = G1 =$
$P(Gender = F|Class = C1) * P(CarType = Sports|Class = C1) * P(AgeGroup =$
$G1|Class = C1) * P(Class = C1)/P = 0.6 * 0.4 * 0.8 * 0.5/P = 0.096/P$

   So the result is Class = C1.

4. $PClass = C0|Gender = F, CarType = Luxury, AgeGroup = G2 =$
$P(Gender = F|Class = C0) * P(CarType = Luxury|Class = C0) * P(AgeGroup =$
$G2|Class = C0) * P(Class = C0)/P = 0.4 * 0.6 * 0.4 * 0.5/P = 0.048/P$

   $PClass = C1|Gender = F, CarType = Luxury, AgeGroup = G2 =$
$P(Gender = F|Class = C1) * P(CarType = Luxury|Class = C1) * P(AgeGroup =$
$G2|Class = C1) * P(Class = C1)/P = 0.6 * 0.6 * 0.2 * 0.5/P = 0.036/P$

   So the result is Class = C0.

2 BN Inference (12 points)

Compute the following probabilities according to the Bayesian net shown in Figure 1. **Note**: $P(A)$ means $P(A = true)$; $P(\sim A)$ means $P(A = false)$.

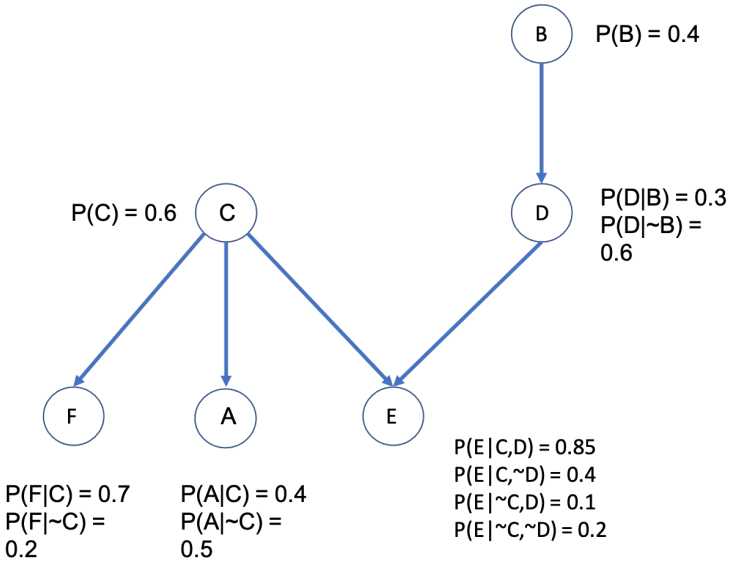

Figure 1: BN Inference

1. Compute $P(A)$. Show your work.
2. Compute $P(D|B, \sim A)$. Show your work.
3. Compute $P(A, B, \sim C, D, E, F)$. Show your work.
4. Are E and F conditionally independent given C? Justify your answer in 1 sentence.
5. Are A and B marginally independent? Justify your answer in 1 sentence.
6. Given evidence that $A = true$, $C = true$ $D = false$, and $F = true$, use the Bayes Net to predict whether E is more likely to be $true$ or $false$, or whether both are equally likely.

1. $P(A) = P(A|C)P(C) + P(A| \sim C)P(\sim C) = 0.4 * 0.6 + 0.5 * 0.4 = 0.44$
2. $P(D|B, \sim A) = P(D|B) = 0.3$
3. $P(A, B, \sim C, D, E, F) = P(A| \sim C)P(B)P(D|B)P(\sim C)P(E| \sim C, D)P(F| \sim C) = 0.5 * 0.4 * 0.3 * 0.4 * 0.1 * 0.2 = 0.00048$
4. T. When C is given precisely, the occurrence of E and the occurrence of F are independent events. If C is unknown, E and F are not independent to each other.
5. T. Knowledge of A's value doesn't affect your belief in the value of B and vice versa.

6. $P(E|A, C, \sim D, F) = 0.4$
   $P(\sim E|A, C, \sim D, F) = P(\sim E|C, \sim D) = 0.6$
   Therefore, E is more likely to be false.