**CSci 384: Artificial Intelligence Spring, 2020**

Derek Trom  **March 31st, 2020**

**Due: by the end of the day, April 7th (Tue.), 2020**

**Home Assignment 5: Uncertainty (100 points)**

**Instruction:**

For each question, do **NOT** verbally describe what & how to compute it **but *define*** what needs to be compute and ***show*** the computational steps precisely.

1. Define the ***variables*** e.g.) A: Alarm, F: Fire, or

use the defined variables: e.g.) E3 and B, etc. in Q2, TA, TB, V, etc. in Q3, etc.

1. Define what needs to be computed: e.g.) P(A|F)
2. Derive the formulas for computation, step by step:

e.g.) P(A|F) = P(F| A)⋅P(A)

1. Assign the values to the derived formula to complete the computation.

e.g.) P(A|F) = P(F|A)⋅P(A) = 0.2 ⋅ 0.6 = 0.12

1. If you write a program to compute any given question, insert the image of output at the corresponding question.

Please read the submission instruction on the blackboard.

If you wrote a program, CREATE a folder first and place all the files in the folder, named ‘HW5-YourLastNameFirstInitial’. Upload the compressed folder on the black board. – The submission of multiple files is NOT appreciated.

**Q1. [20]** There are nine marbles in an urn, three are blue, three pink, and three green. Two marbles are drawn randomly from the urn at the same time in your eyes closed.

1. [10] What is the probability that both marbles are green?
   1. (3 choose 2 \* 3 choose 0 \* 3 choose 0) /(9 choose 2)= 0.0833
2. [10] Suppose you know that both marbles drawn are the same color. What is the probability that both marbles are green?
   1. P(Green|Same) = P(Green & Same) / P(Same)
   2. P(Same) = P(Green & Green) + P(Pink & Pink) + P(Blue & Blue) = 1/12 + 1/12 + 1/12 = ¼
   3. P(Green & Same) = P(Green & Green) = 1/12
   4. So P(Green | Same) = (1/12)/(1/4) = 1/3

**Q2. [10]** A physical exam is administered to all new inmates at a prison. Suppose that 85% of all healthy individuals pass this exam, 70% of all individuals with minor ailments pass, and 30% of all prisoners with serious ailments also pass. Suppose that **30%** of these new prisoners are actually in good health (event **E1**), 50% have minor ailments (**E2**), and **20%** have major health issues (**E3**). Given that an inmate passes this physical (event **B**), what is the *posterior probability that the inmate is in good health*?

Compute it using Bayes’ Theorem, **not** by the enumeration of Full Joint Probability Distribution.

1. H = healthy, M= minor ailments, S = serious ailments, P = physical exam, G = Passed Exam
2. P(G|H) = 0.85, P(G|M) = 0.70, P(G|S) = 0.30,
3. P(H) = 0.30, P(M) = 0.50, P(S) = 0.20
4. Given Baye’s Rule P(a|b) = P(b|a) \* P(a) / P(b)
   1. P(S) = P(G|H)P(H) + P(G|M)P(M) + P(G|S)P(S) = 0.255 + 0.35+0.06=0.665
   2. P(H|G) = P(G|H) \* P(H) / P(G) = (0.85 \* 0.30) / 0.665 = 0.383

**Q3. [25]** Consider two medical tests, A and B, for a virus.

Test A (**TA**) is 95% effective at recognizing the virus (**V**) when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B (**TB**) is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus.

The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus.

Query: Which test returning positive is more indicative of someone really carrying the virus?

Well if this isn’t relevant I don’t know what is

1. [5] Define the variables and their domain that are required in this scenario.
   1. V = virus ∈ {present, absent}
   2. TA = Test A ∈ {positive, negative}
   3. TB = Test B ∈ {positive, negative}
2. [5] Using the defined variables in 1), define the formula(s) of probability that you need to compute to answer the query. e.g.) P(TA|TB).
   1. P(V|TA) and P(V|TB)
   2. P(TA| V) = 0.95
   3. P(TA| ¬V) = 0.10
   4. P(TB|V) = 0.90
   5. P(TB|¬V) = 0.01
3. [10] Compute the probability in 2) using ***Bayes’ Theorem***, **not** by the enumeration of Full Joint Probability Distribution.
   1. P(V|TA) = P(TA|V)P(V)/P(TA) where P(TA) = P(TA & V) +P(TA & ¬V) = 0.087
   2. P(V|TB) = P(TB|V)P(V)/P(TB) where P(TB) = P(TB & V) +P(TB & ¬V) = 0.154
4. [5] What is the answer to query? i.e. which test returning positive is more indicative of someone really carrying the virus? Justify your answer based on your computational result in 3).
   1. TB is more indicative, in the above calculation it has a higher rate .154 compared to .087.

**Q4. [25]** The gardening company wants to statistically analyze his yearly harvest of peas. For every pea pod he picks he measures its length *xi* in centimeters and its weight *yi* in grams. He divides the peas into two classes, the good and the bad (empty pods). The measured data (*xi*, *yi*) are

Good peas:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 2 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 |
| y | 2 | 3 | 4 | 4 | 4 | 5 | 6 | 5 | 6 | 6 |

Bad peas:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | 3 | 4 | 6 | 6 | 7 |
| y | 1 | 2 | 2 | 3 | 3 |

1. Info
   1. G = Good, B = Bad, Yi= Weight, Xi =Length
2. [5] Compute the probability of the peas that are ≥ 3 grams in good peas.
   1. P(G) = P(G & Pea) /P(Pea) = 10/15
   2. P(Yi≥3|G) = P(Yi≥3 & G)/P(G)= (9/15)/(10/15) = 9/10
3. [5] Compute the probability of the peas that are ≥ 3 grams.
   1. P(Pea) = P(G) + P(B) = 15/15
   2. P(Yi ≥ 3|Pea) = P(Pea & Yi ≥ 3) / P(Pea) = (11/15)/(15/15) = 11/15
4. [5] Compute the probability of the good peas.
   1. P(G | Pea) = P(G & Pea) / P(Pea) = (10/15) / (15/15) = 2/3
5. [5] Compute the probability of the good peas in the peas which are ≥ 3 grams.
   1. P(G| Yi≥3) = P(G& Yi≥3) / P(Pea & Yi≥3) = (9/15)/(11/15) = 9/11
6. [5] Compute the probability of the good peas in the peas which are < 3 grams.
   1. P(G| Yi<3) = P(G & Yi<3) / P(Pea & Yi<3) = (1/15) / (4/15) = 1/4

**Q5. [20]** You are supposed to predict the afternoon weather using a few simple weather values from the morning of this day. The classical probability calculation for this requires a complete model, which is given in the following table, where

|  |  |  |  |
| --- | --- | --- | --- |
| ***Sky*** | ***Bar*** | ***Prec*** | ***P(Sky, Bar, Prec)*** |
| Cloudy | Rising | Dry | 0.4 |
| Cloudy | Rising | Raining | 0.08 |
| Cloudy | Falling | Dry | 0.07 |
| Cloudy | Falling | Raining | 0.1 |
| Clear | Rising | Dry | 0.09 |
| Clear | Rising | Raining | 0.11 |
| Clear | Falling | Dry | 0.03 |

*Sky*: The sky is clear or cloudy in the morning.

*Bar*: Barometer rising or falling in the morning

*Prec*: Raining or dry in the afternoon.

1. [10] How many events are in the distribution for these three variables?
   1. 2 choices for each of the 3 variables 23 = 8 distributions for the three variables.
2. [10] Compute P(Prec = dry | Sky = clear, Bar = rising).
   1. P(Prec = dry | Sky = clear, Bar = rising) = P(Sky = clear, Bar = Rising, Prec = dry)/P(Sky=clear, Bar = rising)
   2. = .09 / (.09 + .11) = 0.45
3. Compute P(Prec = rain | Sky = cloudy).
   1. P(Prec = rain | Sky = cloudy) = P(Prec = rain & Sky = Cloudy) / P(Sky = cloudy)
   2. = (0.1 + 0.08) /(0.4 + 0.08 + 0.07 + 0.1) = 0.18 /0.65 = 0.28
4. What would you do if the last row were missing from the table?
   1. Including the already one missing row below
   2. P(Sky = Clear, Bar = Falling, Prec = Raining) = 1 –ΣP(Sky, Bar, Prec)= 0.12
   3. Add the removed row probability to already missing probability = .03+ .12 = .15
   4. Then divide by two 0.15/2 = 0.075 using the indifference principle assign the two missing rows with the 0.075 in order for them to be symmetric.