CS440

Homework 7 (due Tue. Nov. 28th)

1. **[15 points]**
   Show that the statement $P(A, B|C) = P(A|C)P(B|C)$ is equivalent to either of the statements $P(A|B, C) = P(A|C)$ or $P(B|A, C) = P(B|C)$.

2. **[20 points]**
   Suppose we wish to compute $P(H|E_1, E_2)$ for variables H, E_1, E_2, and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?

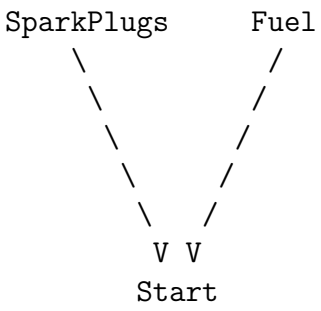
   (a) $P(E1, E2)$, $P(H)$, $P(E_1|h)$, $P(E_2|H)$.

   (b) $P(E_1, E_2)$, $P(H)$, $P(E_1, E_2|H)$

   (c) $P(H)$, $P(E_1, H)$, $P(E_2|H)$.

   Suppose we know that $P(E_1|H, E_2) = P(E_1|H)$ for all values of H, E_1, E_2. Which of the three sets are sufficient?

3. **Explaining away [20 pts]**
   In class we considered a Bayesian network for the car start problem. Part of the network is as follows:



   The two causes `SparkPlugs` and `Fuel` "compete" to explain the observed data. We noted that although marginally independent, they become dependent given that their common child, `Start`, is observed. For example, suppose the car does not start and we know that we are out of fuel, then the probability that the spark plugs are faulty goes down. This phenomenon is called "explaining away". Show that

$$P(SparkPlugs = False|Start = False, Fuel = False) \leq P(SparkPlugs = False|Start = False).$$

   Assume that all entries in the conditional probability tables are nonzero.

4. **[20 pts]**
   Question 14.3 (a) and (b) in Russel and Norvig.

5. **[25 pts]**
   You have noticed that people that drive SUVs (S) consume large amounts of gas (G) and are involved in more accidents than the national average (A). You constructed the following Bayesian network:

   ```
   A <--- S ---> G
   ```

   The conditional probability tables for the networks are:

$P(S = True)$
0.3

S	$P(A = True)$
True	0.6
False	0.3

S	$P(G = True)$
True	0.7
False	0.2

   - Compute P(A) in two ways:
     - By generating the entire joint distribution over these variables and explicitly summing the appropriate entries.
     - Using the variable elimination algorithm. You can check yourself by entering the network into the *tetrad* program which is available from `http://www.phil.cmu.edu/projects/tetrad/tetrad4.html`.
   - You also notice that there are two types of people that drive SUVs, people from Colorado (C) and people with large families (F). How would you modify the network to incorporate this information?

Submission This homework is optional, and if you submit it, it can only improve your homework grade. Due in class on the due date.