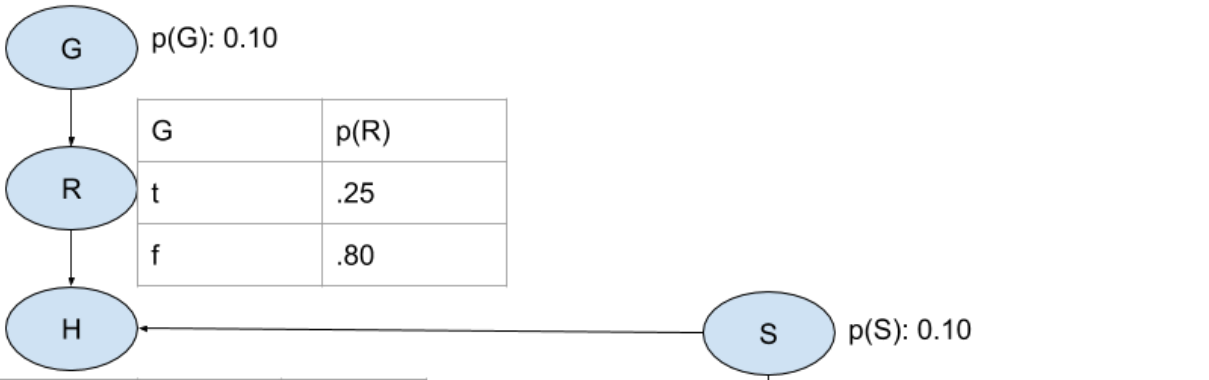
Erikson Sodergren
Intro to IS: Homework 4

1.
   a. p(caff)=16/30=0.53
   b. p(hot|sweet)= (3/30) / (13/30) = 0.10 / 0.43 = 0.23
   c. If p(A&B)=p(A)*p(B), independent.
      i. p(hot)=0.47, p(caff)=0.53, p(sweet)=0.43
      ii. p(hot&caff)=0.13, p(hot&sweet)=0.10, p(caff&sweet)=0.30
      iii. p(hot)*p(caff) = 0.25, p(hot)* p(sweet) = 0.20,p(caff)* p(sweet) = 0.23
      iv. None are independent



G	p(R)
t	.25
f	.80

R	S	p(H)
t	t	.90
t	f	.80
f	t	.30
f	f	.20

S	p(O)
t	.90
f	.50

2.
   a. P(r) = .25*.10 + .80*.90 = 0.745
   b. P(g ∧ s) = .10 * .10 = .01
   c. P(¬h) = 1-(p(h)) = (.745*.10*.10)+(.745*.9*.2)+(.255*.1*.7)+(.255*.9*.8) = .343
   d. P(r ∧ ¬h) = (.745*.1*.1)+(.745*.9*.2) = .14155
   e. P(s | o) = p(o |s)*p(s)/p(o) = .9*.1/(.9*.1+.5*.9) = .167
3.
   a. Predicting a student's grade on the second midterm: grade on the first midterm
      i. grade on the first midterm: useful and practical. Students tend to get similar grades.
      ii. grade in CS2: not very useful. Success in a low level class doesnt mean much for a high level class.
      iii. amount of time studying: useful but impractical. Would rely on self-reporting which is usually innacurate

iv. number of characters on the student's cheat sheet: not useful. A student might write little due to knowing the content, or a lot just as a reminder just in case.
v. amount of coffee consumed in the last week: not useful. Seems largely unrelated

b. Predicting whether it will rain in Rochester tomorrow
i. whether it rained in Rochester today: useful and practical. Less likely to rain 2 days in a row.
ii. whether it rained in Cleveland today: not useful without doing global weather calculations
iii. all 100-square-meter areas in the US where it rained yesterday: useful at least for nearby locations, but very impractical.
iv. the wind speed in Rochester yesterday: useful and practical. Certain windspeeds could possibly indicate coming rain.
v. the day of the week: not useful
vi. the month of the year: useful and practical. It won't rain in december, it will snow….we hope

c. Predicting the score of RIT's next hockey game
i. the average heights of the players on RIT's team and the opponent's team: probably useful, definitely practical. Every sport has body proportions that are most beneficial.
ii. the temperature outside at game time: not useful. Outdoor conditions have no impact on indoor events.
iii. the opponent's win/loss record this season: useful and practical. A team that wins a lot already is more likely to win this time.
iv. the number of goals scored per minute of each player on both teams so far this season: useful and practical. Similar to iii
v. the number of goals scored per kilometer skated by each player so far this season: useful for same reasons as iii and iv, but impractical.

d. Predicting whether you will like a restaurant
i. the opinions of the last hundred people to eat there: useful but impractical. If lots of people have a strong opinion it might be for a good reason, but getting those 100 people to answer is difficult.
ii. the number of stars in the Yelp review of the restaurant: not useful. Yelp practically makes you pay to have a good rating.
iii. the type of food: useful and practical. You won't like a place that sells mostly food you hate.
iv. the number of insects in the kitchen: useful but impractical. A buggy place is worse, but how do you count them?

4. Using 0 based indexing for the attributes
a. if entry[3] <= 0:
b. if entry[4] <= 0:
c. return 0

```
        else:
            return 0
    else:
        if entry[5] <= 0:
            return 1
        else:
            return 0
```

5. Stump: return 0 if entry[3] = 0, else return 1
    a. 35/200 wrong answers for this stump, error = 0.175
    b. .6734
    c. 1/200 = .005
    d. Correct: .00106, incorrect: .005 before normalization(sum of all weights=.3499)
        i. Correct: .00303, incorrect: .01429 after normalization