Machine Learning

Assignment #5

Due: Monday, March 13th

1. Answer the following:
    a) The word "offer" appears in 80% of spam messages in my email account. The word "offer" also occurs in 10% of my desired email. If 30% of my overall email is spam and I receive a new message that contains the word "offer", what is the probability that it's spam?

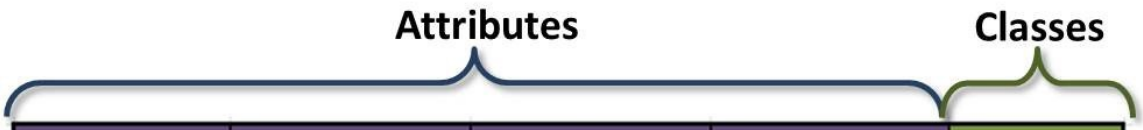
    b) An art competition has entries from only three painters: Pam, Pia and Pablo. Pam submitted 15 paintings, and 4% of her works have won first prize in past shows. Pia submitted 5 paintings, and 6% of her works have won first prize in past shows. Pablo submitted 10 paintings, and 3% of his works have won first prize in past shows. What is the probability that Pam will win first prize?

2. Determine the MAP hypotheses for the following:
    a) A radar system accurately detects the presence of aircraft in its range 98% of the time. However, if no aircraft is present it still reports (falsely) that one is present 5% of the time. At any given moment, the probability that an aircraft is present within the range of the radar is 7%. Is there an aircraft in the radar's space right now?

    b) Marie is getting married tomorrow at an outdoor ceremony in the desert. Unfortunately, the weatherman has predicted rain for tomorrow. In recent times, it has rained 7 days each year. When it actually rains, the weatherman correctly forecasts it 80% of the time. When it doesn't rain, he incorrectly forecasts it 10% of the time. Should Marie expect rain on not?

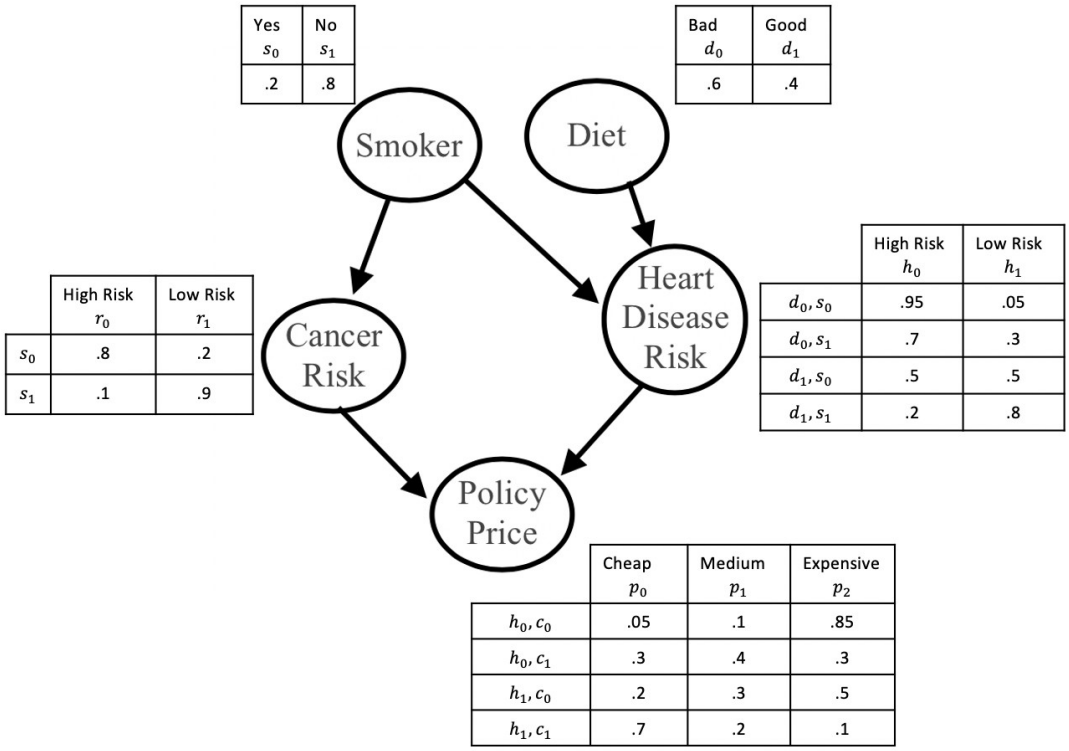
3. Given the dataset shown below, classify the following datum using a Naïve Bayesian Classifier:

a. X=(Outlook=Rainy, Temperature=Mild, Humidity=Normal, Windy=True)
b. X=(Outlook=Sunny, Temperature=Cool, Humidity=High, Windy=False)
c. X=(Outlook=Overcast, Temperature=Hot, Humidity=Normal, Windy=False)



Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No

4. Given the Bayesian Network shown below, calculate the following probabilities:
   a. (Policy=Cheap, Cancer Risk=High, Heart Risk=Low, Smoker=Yes, Diet=Good)
   b. (Policy=Expensive, Cancer Risk=Low, Heart Risk=High, Smoker=No, Diet=Bad)
   c. (Policy=Medium, Cancer Risk=Low, Heart Risk=Low, Smoker=Yes, Diet=Bad)



Yes s_0	No s_1
.2	.8

Bad d_0	Good d_1
.6	.4

	High Risk r_0	Low Risk r_1
s_0	.8	.2
s_1	.1	.9

	High Risk h_0	Low Risk h_1
d_0, s_0	.95	.05
d_0, s_1	.7	.3
d_1, s_0	.5	.5
d_1, s_1	.2	.8

	Cheap p_0	Medium p_1	Expensive p_2
h_0, c_0	.05	.1	.85
h_0, c_1	.3	.4	.3
h_1, c_0	.2	.3	.5
h_1, c_1	.7	.2	.1