Team Members:

Shanmukha Datta Sai Sasank Tadepalli - NetID: na2500
Bhavana Kamasani - NetID: tl1593
Neeraj Kumar Maraka - NetID: ho4913

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?

Answer:

a)

1)

a) Given,

$$P(\text{offer} \mid \text{Spam}) = 0.8$$
$$P(\text{offer} \mid \text{Not Spam}) = 0.1$$
$$P(\text{Spam}) = 0.3$$

$$P(\text{offer}) = P(\text{offer} \mid \text{Spam}) * P(\text{Spam}) + P(\text{offer} \mid \text{Not Spam}) * P(\text{Not Spam})$$
$$= 0.8 * 0.3 + 0.1 * 0.7$$
$$= 0.24 + 0.07$$
$$= 0.31$$

Now, $$P(\text{Spam} \mid \text{offer}) = \frac{P(\text{offer} \mid \text{Spam}) * P(\text{Spam})}{P(\text{offer})}$$
$$= \frac{0.8 * 0.3}{0.31}$$
$$= 0.77419355$$

Therefore, the probability that an email containing “offer” is spam is 0.77419355 b)

b) Given,

$P(\text{First} \mid \text{Pam}) = 0.04$
$P(\text{First} \mid \text{Pia}) = 0.06$
$P(\text{First} \mid \text{Pablo}) = 0.03$
Number of Paintings by Pam = 15
Number of Paintings by Pia = 5
Number of Paintings by Pablo = 10

Total Number of Paintings = 15 + 5 + 10 = 30

$P(\text{Pam}) = 15/30 = 1/2$
$P(\text{Pia}) = 5/30 = 1/6$
$P(\text{Pablo}) = 10/30 = 1/3$

$$P(\text{First}) = P(\text{Pam}) * P(\text{First} \mid \text{Pam}) + P(\text{Pia}) * P(\text{First} \mid \text{Pia}) + P(\text{Pablo}) * P(\text{First} \mid \text{Pablo})$$
$$= 1/2 * 0.04 + 1/6 * 0.06 + 1/3 * 0.03$$
$$= 0.02 + 0.01 + 0.01$$
$$= 0.04$$

Now, $$P(\text{Pam} \mid \text{First}) = \frac{P(\text{First} \mid \text{Pam}) * P(\text{Pam})}{P(\text{First})}$$
$$= \frac{0.04 * 1/2}{0.04}$$
$$= 0.5$$

Therefore, the probability that Pam will win first prize is 0.5

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?

Answer:

a)

a) Given,

$P(\text{Detect} \mid \text{Aircraft}) = 0.98$

$P(\text{Detect} \mid \text{No Aircraft}) = 0.05$

$P(\text{Aircraft}) = 0.07$

So, $P(\text{No Aircraft}) = 1 - P(\text{Aircraft})$

$= 1 - 0.07$

$= 0.93$

Now,

$P(\text{Aircraft} \mid \text{Detect}) = P(\text{Detect} \mid \text{Aircraft}) * P(\text{Aircraft})$

$= 0.98 * 0.07$

$= 0.0686$

$P(\text{No Aircraft} \mid \text{Detect}) = P(\text{Detect} \mid \text{No Aircraft}) * P(\text{No Aircraft})$

$= 0.05 * 0.93$

$= 0.0465$

Since, P(Aircraft | Detect) > P(No Aircraft | Detect)

h_{MAP} = Aircraft

Therefore, there is an Aircraft in the radar's space right now.

b)

b) Given,

$P(\text{Rain}) = 7/365 = 0.01917808$

$P(\text{Forecasted} \mid \text{Rain}) = 0.8$

$P(\text{Forecasted} \mid \text{No Rain}) = 0.1$

So, $P(\text{No Rain}) = 1 - P(\text{Rain}) = 1 - 0.01917808$

$= 0.98082192$

Now,

$P(\text{Rain} \mid \text{Forecasted}) = P(\text{Forecasted} \mid \text{Rain}) * P(\text{Rain})$

$= 0.8 * 0.01917808$

$= 0.015342464$

$P(\text{No Rain} \mid \text{Forecasted}) = P(\text{Forecasted} \mid \text{No Rain}) * P(\text{No Rain})$

$= 0.1 * 0.98082192$

$= 0.098082192$

Since, P(No Rain | Forecasted) > P(Rain | Forecasted)
h_{MAP} = No Rain
Therefore, there won't be any rain tomorrow.

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)

Attributes				Classes
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

Answer:
Some required calculations from the given data:

$P(Yes) = 9/14 = 0.643$
$P(Windy = True \mid Yes) = 3/9 = 0.333$
$P(Windy = False \mid Yes) = 6/9 = 0.667$
$P(Humidity = High \mid Yes) = 3/9 = 0.333$
$P(Humidity = Normal \mid Yes) = 6/9 = 0.667$
$P(Temperature = Hot \mid Yes) = 2/9 = 0.222$
$P(Temperature = Mild \mid Yes) = 4/9 = 0.444$
$P(Temperature = Cool \mid Yes) = 3/9 = 0.333$
$P(Outlook = Sunny \mid Yes) = 3/9 = 0.333$
$P(Outlook = Overcast \mid Yes) = 4/9 = 0.444$
$P(Outlook = Rainy \mid Yes) = 2/9 = 0.222$

$P(No) = 5/14 = 0.357$
$P(Windy = True \mid No) = 3/5 = 0.6$
$P(Windy = False \mid No) = 2/5 = 0.4$
$P(Humidity = High \mid No) = 4/5 = 0.8$
$P(Humidity = Normal \mid No) = 1/5 = 0.2$
$P(Temperature = Hot \mid No) = 2/5 = 0.4$
$P(Temperature = Mild \mid No) = 2/5 = 0.4$
$P(Temperature = Cool \mid No) = 1/5 = 0.2$
$P(Outlook = Sunny \mid No) = 2/5 = 0.4$
$P(Outlook = Overcast \mid No) = 0/5 = 0$
$P(Outlook = Rainy \mid No) = 3/5 = 0.6$

a) X=(Outlook=Rainy, Temperature=Mild, Humidity=Normal, Windy=True)

a) $X = (Outlook = Rainy, Temperature = Mild, Humidity = Normal, Windy = True)$

$$P(Yes \mid X) = P(Yes) * P(Outlook = Rainy \mid Yes) * P(Temperature = Mild \mid Yes) * P(Humidity = Normal \mid Yes) * P(Windy = True \mid Yes)$$
$$= 0.643 * 0.222 * 0.444 * 0.667 * 0.333$$
$$= 0.01407$$

$$P(No \mid X) = P(No) * P(Outlook = Rainy \mid No) * P(Temperature = Mild \mid No) * P(Humidity = Normal \mid No) * P(Windy = True \mid No)$$
$$= 0.357 * 0.6 * 0.4 * 0.2 * 0.6$$
$$= 0.01028$$

Since, probability of Yes > No, we classify the given data as a YES
Therefore, if Outlook=Rainy, Temperature=Mild, Humidity=Normal, Windy=True we can say YES to Playing Golf.

b) X=(Outlook=Sunny, Temperature=Cool, Humidity=High, Windy=False)

b) $X = (\text{Outlook} = \text{Sunny}, \text{Temperature} = \text{Cool}, \text{Humidity} = \text{High}, \text{Windy} = \text{False})$

$$P(\text{Yes} \mid X) = P(\text{Yes}) * P(\text{Outlook} = \text{Sunny} \mid \text{Yes}) * P(\text{Temperature} = \text{Cool} \mid \text{Yes}) * P(\text{Humidity} = \text{High} \mid \text{Yes}) * P(\text{Windy} = \text{False} \mid \text{Yes})$$

$$= 0.643 * 0.333 * 0.333 * 0.333 * 0.667$$

$$= 0.01583$$

$$P(\text{No} \mid X) = P(\text{No}) * P(\text{Outlook} = \text{Sunny} \mid \text{No}) * P(\text{Temperature} = \text{Cool} \mid \text{No}) * P(\text{Humidity} = \text{High} \mid \text{No}) * P(\text{Windy} = \text{False} \mid \text{No})$$

$$= 0.357 * 0.4 * 0.2 * 0.8 * 0.4$$

$$= 0.00913$$

Since, probability of Yes > No, we classify the given data as a YES
Therefore, if Outlook=Sunny, Temperature=Cool, Humidity=High, Windy=False we can say YES to Playing Golf.

c) X=(Outlook=Overcast, Temperature=Hot, Humidity=Normal, Windy=False)

c) $X = (\text{Outlook} = \text{Overcast}, \text{Temperature} = \text{Hot}, \text{Humidity} = \text{Normal}, \text{Windy} = \text{False})$

$$P(\text{Yes} \mid X) = P(\text{Yes}) * P(\text{Outlook} = \text{Overcast} \mid \text{Yes}) * P(\text{Temperature} = \text{Hot} \mid \text{Yes}) * P(\text{Humidity} = \text{Normal} \mid \text{Yes}) * P(\text{Windy} = \text{False} \mid \text{Yes})$$

$$= 0.643 * 0.444 * 0.222 * 0.667 * 0.667$$

$$= 0.02819$$

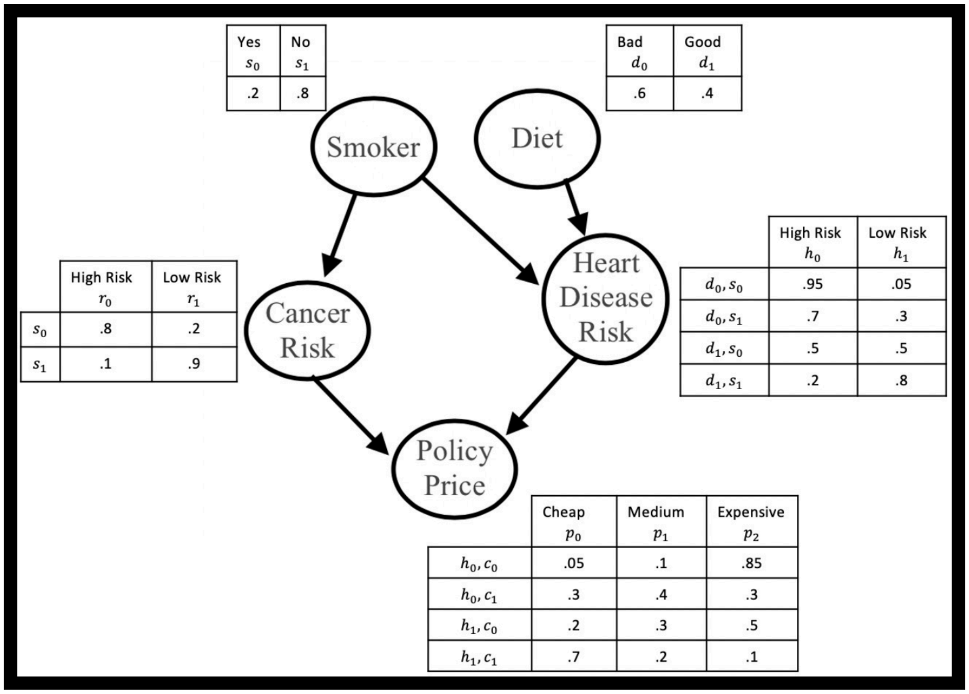
$$P(\text{No} \mid X) = P(\text{No}) * P(\text{Outlook} = \text{overcast} \mid \text{No}) * P(\text{Temperature} = \text{Hot} \mid \text{No}) * P(\text{Humidity} = \text{Normal} \mid \text{No}) * P(\text{Windy} = \text{False} \mid \text{No})$$

$$= 0.357 * 0 * 0.4 * 0.2 * 0.4$$

$$= 0$$

Since, probability of Yes > No, we classify the given data as a YES
Therefore, if Outlook=Overcast, Temperature=Hot, Humidity=Normal, Windy=False we can say YES to Playing Golf.

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)



Answer:

a) (Policy=Cheap, Cancer Risk=High, Heart Risk=Low, Smoker=Yes, Diet=Good)

a) P(Policy = Cheap, Cancer Risk = High, Heart Risk = low, Smoker = Yes, Diet = Good)

$= P(P = P_0, C = C_0, H = h_1, S = S_0, D = d_1)$

$= P(P = P_0 \mid C = C_0, H = h_1) * P(C = C_0 \mid S = S_0) * P(H = h_1 \mid S = S_0, D = d_1) * P(S = S_0) * P(D = d_1)$

$= 0.2 * 0.8 * 0.5 * 0.2 * 0.4$

$= 0.0064$

Therefore Probability that Policy=Cheap, Cancer Risk=High, Heart Risk=Low, Smoker=Yes, Diet=Good is 0.0064

b) (Policy=Expensive, Cancer Risk=Low, Heart Risk=High, Smoker=No, Diet=Bad)

b) $P(\text{Policy} = \text{Expensive}, \text{Cancer Risk} = \text{Low}, \text{Heart Risk} = \text{High}, \text{Smoker} = \text{No}, \text{Diet} = \text{Bad})$

$$= P(P=P_2, C=C_1, H=h_0, S=S_1, D=d_0)$$

$$= P(P=P_2 \mid C=C_1, H=h_0) * P(C=c_1 \mid S=S_1) * P(H=h_0 \mid S=S_1, D=d_0) * P(S=S_1) * P(D=d_0)$$

$$= 0.3 * 0.9 * 0.7 * 0.8 * 0.6$$

$$= 0.09072$$

Therefore Probability that Policy=Expensive, Cancer Risk=Low, Heart Risk=High, Smoker=No, Diet=Bad is 0.09072

c) (Policy=Medium, Cancer Risk=Low, Heart Risk=Low, Smoker=Yes, Diet=Bad)

c) $P(\text{Policy} = \text{Medium}, \text{Cancer Risk} = \text{Low}, \text{Heart Risk} = \text{Low}, \text{Smoker} = \text{Yes}, \text{Diet} = \text{Bad})$

$$= P(P=P_1, C=C_1, H=h_1, S=S_0, D=d_0)$$

$$= P(P=P_1 \mid C=C_1, H=h_1) * P(C=c_1 \mid S=S_0) * P(H=h_1 \mid S=S_0, D=d_0) * P(S=S_0) * P(D=d_0)$$

$$= 0.2 * 0.2 * 0.05 * 0.2 * 0.6$$

$$= 0.00024$$

Therefore Probability that Policy=Medium, Cancer Risk=Low, Heart Risk=Low, Smoker=Yes, Diet=Bad is 0.00024