# Exercise 4

Nitai Aharoni - 203626742

**Part 1 - Probability Questions**

For each one of the following question please supply your calculations alongside your answers.

1. There are 2 boxes. The first has 3 green balls and 9 red balls. The second one has 5 green balls and 7 red balls. One ball is transferred from the first box to the second box. Now, A ball is drawn from the second box. Given that the drawn ball is green, find the probability that the transferred ball is red.

**.1**

A – הכדור שהועבר מהקופסא הראשונה לשנייה הוא אדום

B – הכדור שהוצא מהקופסא השנייה הוא ירוק

- נחשב את P(B) לפי הסתברות שלמה: $P(B) = P(B|A) * P(A) + P(B|A') * P(A') = \frac{5}{13} * \frac{9}{12} + \frac{6}{13} * \frac{3}{12} = \frac{21}{52}$
ההסתברות שהוצא כדור ירוק אם עבר כדור ירוק + ההסתברות שהוצא כדור ירוק אם עבר כדור אדום

- נחשב את $P(A \cap B)$: $P(A \cap B) = \frac{9}{12} * \frac{5}{13} = \frac{15}{52}$
ההסתברות שהוצא כדור ירוק וגם הועבר כדור אדום

- נחשב את P(A|B): $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{15}{52}}{\frac{21}{52}} = \frac{15}{21} = \mathbf{0.7142}$
ההסתברות שהועבר כדור אדום בהינתן שהוצא כדור ירוק

2. You and a friend are going to Shilat tomorrow for a day of climbing. In recent years, the conditions were good for climbing only 60 days each year. Dani Rup (the weatherman) has predicted good climbing conditions for tomorrow. When the climbing conditions are actually good, Dani correctly forecasts the conditions in 85% of the time. When the climbing conditions aren't good, he incorrectly forecasts good conditions 10% of the time. What is the probability that the climbing condition will be good tomorrow?

A – התנאים טובים

B – החזאי חזה שיהיו תנאים טובים

- נתון P(A): $P(A) = \frac{60}{365} = \frac{12}{73}$
- נתון $P(B|A) = 0.85$ – בהינתן תנאים טובים, ההסתברות שהחזאי חוזה שיהיו תנאים טובים היא 0.85
- נתון $P(B|A') = 0.1$ – בהינתן תנאים לא טובים, ההסתברות שהחזאי חוזה שיהיו תנאים טובים היא 0.1
- נחשב את P(B) הסתברות שלמה: $P(B) = P(A) * P(B|A) + P(A') * P(B|A') = \frac{12}{73} * 0.85 + \frac{61}{73} * 0.1 = 0.2232$
- נחשב את P(A|B): $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B|A)}{P(B)} = \frac{\frac{12}{73} * 0.85}{0.2232} = \mathbf{0.6257}$
בהינתן שהחזאי חזה שיהיו תנאים טובים, ההסתברות שיהיו תנאים טובים

3. You toss 3 fair dice independently. Every die has 4 possible results (1, 2, 3, 4). Let A denote the event that the sum of the dice is 7. Let B denote the event that the sum of the dice is divisible by 3. Let X be a random variable denoting the outcome of the toss of the first die.

A – סכום המספרים על הקוביות הוא 7

B – סכום המספרים על הקוביות מתחלק ב-3

X – התוצאה של הטלת הקובייה הראשונה

(a) Compute P (A|X = i) for each i = {1, 2, 3, 4}

- $P(X=i)=\frac{1}{4}$ – הסתברות של קוביה הוגנת בעלת 4 פאות
- סך הכל תוצאות אפשריות להטלות 3 קוביות הן: $4^3=64$
- $P(A|X=1)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{2,4},{4,2},{3,3}} – כלומר 3 אפשרויות
  - לכן: $P(A|X=1)=\frac{P(A\cap X=1)}{P(X=1)}=\frac{\frac{3}{64}}{\frac{1}{4}}=0.1875$
- $(A|X=2)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{2,3},{3,2},{1,4},{4,1}}– כלומר 4 אפשרויות
  - לכן: $P(A|X=1)=\frac{P(A\cap X=1)}{P(X=1)}=\frac{\frac{4}{64}}{\frac{1}{4}}=0.25$
- $P(A|X=3)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{1,3},{3,1},{2,2}} – כלומר 3 אפשרויות
  - לכן: $P(A|X=1)=\frac{P(A\cap X=1)}{P(X=1)}=\frac{\frac{3}{64}}{\frac{1}{4}}=0.1875$
- $P(A|X=4)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{1,2},{2,1}} – כלומר 2 אפשרויות
  - לכן: $P(A|X=1)=\frac{P(A\cap X=1)}{P(X=1)}=\frac{\frac{2}{64}}{\frac{1}{4}}=0.125$

(b) Compute P (A)

$$\begin{aligned}P(A) &= P(A|X=1)*P(X=1)+P(A|X=2)*P(X=2)+P(A|X=3)*P(X=3)+P(A|X=4)*P(X=4)\\ &= 0.25*\big(P(A|X=1)+P(A|X=2)+P(A|X=3)+P(A|X=4)\big)\\ &= 0.25*(0.1875+0.1875+0.25+0.125)=\mathbf{0.1875}\end{aligned}$$

(c) Are the event A and X = i independent for each i = {1, 2, 3, 4}

- עבור X=1 כן ב"ת: $P(A)*P(X=1)=0.1875*0.25=0.0468=0.0468=\frac{3}{64}=P(A\cap X=1)$
- עבור X=2 לא ב"ת: $P(A)*P(X=2)=0.1875*0.25=0.0468\neq 0.0625=\frac{4}{64}=P(A\cap X=2)$

- עבור X=3 כן ב"ת: $P(A\cap X=3)=\frac{3}{64}=0.0468=0.0468=0.1875*0.25=P(A)*P(X=3)$
- עבור X=4 לא ב"ת: $P(A\cap X=4)=\frac{2}{64}=0.0312\neq 0.0468=0.1875*0.25=P(A)*P(X=4)$
- לכן המאורעות תלויים

(d) Compute P(B|X = i) for each i = {1, 2, 3, 4}

- $P(X=i)=\frac{1}{4}$ – הסתברות של קובייה הוגנת בעלת 4 פאות
- סך הכל תוצאות אפשריות להטלות 3 קוביות הן: $4^3=64$
- $P(B|X=1)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יתחלק ב3 הן {{1,4},{4,1},{4,4},{3,2},{2,3},{1,1}} – כלומר 4 אפשרויות
  - לכן: $P(B|X=1)=\frac{P(B\cap X=1)}{P(X=1)}=\frac{\frac{6}{64}}{\frac{1}{4}}=0.375$
- $P(B|X=2)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יתחלק ב3 הן {{1,3},{3,1},{3,4},{4,3},{2,2}} – כלומר 5 אפשרויות
  - לכן: $P(B|X=2)=\frac{P(B\cap X=2)}{P(X=2)}=\frac{\frac{5}{64}}{\frac{1}{4}}=0.3125$
- $P(A|X=3)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{2,4},{4,2},{3,3},{2,1},{1,2}} – כלומר 5 אפשרויות
  - לכן: $P(B|X=3)=\frac{P(B\cap X=3)}{P(X=3)}=\frac{\frac{5}{64}}{\frac{1}{4}}=0.3125$
- $P(A|X=4)$:
  - התוצאות האפשריות של שאר הקוביות כדי שסכום המספרים יהיה 7 הן {{3,2},{2,3},{1,4},{4,1},{4,4},{1,1}} – כלומר 6 אפשרויות
  - לכן: $P(B|X=4)=\frac{P(B\cap X=4)}{P(X=4)}=\frac{\frac{6}{64}}{\frac{1}{4}}=0.375$

(e) Compute P(B)

$$\begin{aligned}P(B)&=P(B|X=1)*P(X=1)+P(B|X=2)*P(X=2)+P(B|X=3)*P(X=3)+P(B|X=4)*P(X=4)\\&=0.25*\left(P(B|X=1)+P(B|X=2)+P(B|X=3)+P(B|X=4)\right)\\&=0.25*(0.375+0.3125+0.3125+0.375)=\mathbf{0.34375}\end{aligned}$$

(f) Are the event B And X = i independent for each i = {1, 2, 3, 4}

- עבור X=1 לא ב"ת: $P(B\cap X=1)=\frac{6}{64}=0.0937\neq 0.0859=0.34375*0.25=P(B)*P(X=1)$

- עבור X=2 לא ב"ת: $P(B \cap X = 2) = \frac{5}{64} = 0.0781 \neq 0.0859 = 0.34375 * 0.25 = P(B) * P(X = 2)$
- עבור X=3 לא ב"ת: $P(B \cap X = 3) = \frac{5}{64} = 0.0781 \neq 0.0859 = 0.34375 * 0.25 = P(B) * P(X = 3)$
- עבור X=4 לא ב"ת: $P(B \cap X = 4) = \frac{6}{64} = 0.0937 \neq 0.0859 = 0.34375 * 0.25 = P(B) * P(X = 4)$
- לכן המאורעות תלויים

4. There are 1000 dice, which look identical. However, 999 of them are" fair" (i.e. when rolling the die, P(X = k) =1/6 and one of them is forged (specifically when rolling that die the probability of getting 1 is 0.9 and 0.02 for the rest). Assuming we selected a die randomly, rolled it ten times, and saw 8 "1" and 2 "6" (you don't know the sequence of events), what is the probability we choose the forged die?

A – יצאה קובייה הוגנת

A' – יצאה הקובייה הלא הוגנת

B – יצא "1" ב8 הטלות ו-"6" ב 2 הטלות

- נתון $P(A) = \frac{999}{1000} = 0.999$
- נתון $P(A') = \frac{1}{1000} = 0.001$
- $P(A \cap B') = 0.001 * (0.9)^8 * (0.02)^2 * \binom{10}{2} = 7.7 * 10\hat{} - 6$
- P(B) לפי הסתברות שלמה: $P(B) = P(B|A) * P(A) + P(B|A') * P(A') = 0.999 * \binom{10}{2} * \left(\frac{1}{6}\right)^{10} + 0.001 * \binom{10}{2} * (0.9)^8 * (0.02)^2$
- נחשב את P(A'|B): $P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{0.001 * (0.9)^8 * (0.02)^2 * \binom{10}{2}}{\binom{10}{2} * (0.999 * \left(\frac{1}{6}\right)^{10} + 0.001 * (0.9)^8 * (0.02)^2)} = \mathbf{0.91244}$

5. In an ancient tribe, the chief decided he does not allow a family to have more than 1 boy as this will threaten his position. Thus, each family must stop child birth once a boy is born. As a result, families in this country are of the following type:

   - Boy
   - Girl, Boy
   - Girl, Girl, Boy
   - Girl, Girl, Girl, Boy
   - Girl, Girl, ..., Girl, Boy
   - And so on...

Assuming the chance of having a boy is 1/3, what is the expected value of the number of girls in a family in this tribe.

$A \sim G(\frac{1}{3})$ – מס' ילדים עד וכולל הולדת בן ראשון מבין סדרת ילודה ב"ת עם סיכוי $\frac{1}{3}$

- נחשב את התוחלת של A (מתפלג גיאומטרית) - $E(A) = \frac{1}{p} = \frac{1}{\frac{1}{3}} = 3$
  כלומר 3 ילדים במשפחה
- לכן מספר הבנות במשפחה שנצפה לו במשפחה בשבט הוא: $(3-1) = 2$ בנות

**Part 2 - Naive Bayes Classifier**

1. In a far away land where the climate is very strange, two researchers, Alice and Bob, are looking to film an exotic creature, the Randomamml.

   Since they are both lazy, and they know the probability of spotting a randommal is 0.2, they only go out of the camp when they think there is a good probability of finding a Randomamml. From previous observation they know something about the temperature and humidity in the days the Randomamml was spotted and the days he was not.

   When **spotted** they know:

   - Humidity $\sim N(0, 1)$
   - Temperature $\sim N(0, 1)$

   and when **not spotted** they know:

   - Humidity $\sim N(0.2, 1)$
   - Temperature $\sim N(0.2, 1)$

   This information is sufficient for naive Bob, but Alice wants to be more accurate, so she searched for more data and managed to infer that temperature and humidity follow a bivariate normal distribution in each one of the situations and further more evaluated the covariance of the humidity and temperature for each of the situations.

   When **spotted**:

   - $\text{cov(humidity, temp)} = 0.8$

   When **not spotted**:

   - $\text{cov(humidity, temp)} = -0.8$

   The forecast for tomorrow is:

   - temperature = 1, humidity = 1

   We know that the Randommal is coming out tomorrow (because we can see into the future). For both Bob (using a Naive Bayes classifier) and Alice (using full Bayes classifier with her inferred density function) say whether they manage to spot the Randommal or not.

   as a reminder here is the bivariate normal distribution density function:

$$p(\bar{x}|A_i) = \frac{1}{2\pi \cdot \sqrt{|S|}} \cdot e^{-\frac{1}{2}\cdot(\bar{x}-\bar{\mu_i})^T \cdot S^{-1} \cdot (\bar{x}-\bar{\mu_i})}$$

Where:

- |S| = the determinant of matrix S.
- $S^{-1}$ = The inverse of matrix S
- T is the transpose operator.

* along side your answer provides full descriptive calculations of how you got your results.

** For the inverse matrix calculations, you can use the following website: http://matrix.reshish.com/inverse.php.

A – היצור נצפה

A’ – היצור לא נצפה

$P(A) = 0.2$

נתונים:

- עבור Bob:
  - כש- A:
    - Humidity ~ N (0, 1)
    - Temperature ~ N (0, 1)
  - כש- A’:
    - Humidity ~ N (0.2, 1)
    - Temperature ~ N (0.2, 1)
- עבור Alice:
  - כש- A:
    - cov(humidity, temp) = 0.8
  - כש- A’:
    - cov(humidity, temp) = -0.8

ידוע ש:

- מזג האוויר למחר הוא: temperature=1, humidity=1

- שהיצור יוצא לשטח מחר

נבדוק האם Bob צפוי להצליח לצפות ביצור:

- $P(A|humidity = 1, temperature = 1) = P(A) * P(humidity = 1|A) * P(temperature = 1|A) = 0.2 * \left(\left(\frac{1}{\sqrt{2\pi}}\right) * e^{-\frac{(1-0)^2}{2}}\right) * \left(\left(\frac{1}{\sqrt{2\pi}}\right) * e^{-\frac{(1-0)^2}{2}}\right) = \frac{0.2}{2\pi e} = 0.01171$
- $P(A'|humidity = 1, temperature = 1) = P(A') * P(humidity = 1|A') * P(temperature = 1|A') = 0.8 * \left(\left(\frac{1}{\sqrt{2\pi}}\right) * e^{-\frac{(1-0.2)^2}{2}}\right) * \left(\left(\frac{1}{\sqrt{2\pi}}\right) * e^{-\frac{(1-0.2)^2}{2}}\right) = \frac{0.8}{2\pi} * e^{-0.64} = 0.06714$
- $P(A|humidity = 1, temperature = 1) = 0.01171 < 0.06714 = P(A'|humidity = 1, temperature = 1)$
  ולכן Bob צפוי לא להצליח לצפות ביצור

נבדוק האם Alice צפויה להצליח לצפות ביצור:

- מטריצות ה-covariance:
  - $S|A = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$
  - $S^{-1}|A = \begin{pmatrix} 2.77 & -2.22 \\ -2.22 & 2.77 \end{pmatrix}$
  - $S|A' = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}$
  - $S^{-1}|A' = \begin{pmatrix} 2.77 & 2.22 \\ 2.22 & 2.77 \end{pmatrix}$
- $X = (1,1)$
- $|(S|A)| = |(S|A')| = 0.36$
- נחשב את $\mathrm{P}(A|X)$:
  - $\mathrm{P}(X|A) = \frac{1}{2\pi * \sqrt{0.36|}} * e^{-\frac{1}{2}(1,1) * \begin{pmatrix} 2.77 & -2.22 \\ -2.22 & 2.77 \end{pmatrix} * \binom{1}{1}} = \frac{1}{2\pi * \sqrt{0.36|}} * e^{-\frac{1}{2}(0.55,\ 0.55) * \binom{1}{1}} == \frac{1}{2\pi * \sqrt{0.36|}} * e^{-\frac{1}{2} * 1.1} = 0.26526 * 0.57695 = 0.15304$
  - $\mathrm{P}(A|X) = \frac{P(X|A) * P(A)}{P(X)} => P(X|A) * P(A) = 0.15304 * 0.2 = 0.0306$
- נחשב את $\mathrm{P}(A'|X)$:
  - $\mathrm{P}(X|A') = 0.26526 * e^{-\frac{1}{2}(1-0.2, 1-0.2) * \begin{pmatrix} 2.77 & 2.22 \\ 2.22 & 2.77 \end{pmatrix} * \binom{1-0.2}{1-0.2}} = 0.26526 * e^{-\frac{1}{2} * 6.3872} = 0.26526 * 0.04102 = 0.01088$
  - $\mathrm{P}(A'|X) = \frac{P(X|A') * P(A\prime)}{P(X)} => P(X|A') * P(A') = 0.01088 * 0.8 = 0.0087$
- $\mathrm{P}(X|A) = 0.0306 > 0.0087 = \mathrm{P}(X|A')$
  ולכן Alice צפויה כן להצליח לצפות ביצור

2. The owner of the famous Randomistan restaurant comes to seek your help. after 10 themed nights he gathered some data and wants to prepare dishes that are aligned with that data. He has the following information about the dishes:
at what night they were served, were they spicy, the cooking method and whether they were returned or not. He seeks your advice with the following 2 dishes:

(a) A normal grilled dish in a Mexican themed night.

(b) A spicy steamed dish in an Indian themed night.

Here is the gathered dataset:

| **Theme** | **Spiced** | **Cooking** | **Returned** |
|---|---|---|---|
| Mexican | Spicy | Fried | False |
| Mexican | Normal | Grilled | True |
| Mexican | Spicy | Grilled | False |
| Mexican | Normal | Fried | True |
| Mexican | Spicy | Grilled | True |
| Indian | Spicy | Fried | True |
| Indian | Normal | Steamed | False |
| Indian | Spicy | Fried | False |
| Indian | Normal | Grilled | False |
| Indian | Normal | Grilled | False |

Using a Naive Bayes Classifier with Laplace smoothing help the owner decide whether he should serve those dishes or not.

* along side your answer provides full descriptive calculations of how you got your results.

A – מנה הוחזרה

A’ – מנה לא הוחזרה

- $P(A) = \frac{4}{10} = \frac{2}{5}$
- $P(A) = \frac{6}{10} = \frac{3}{5}$
- עבור A (Laplance):
    - Theme:
        - $\text{P}(Indian|A) = \frac{1+1}{4+2} = \frac{2}{6} = \frac{1}{3}$
        - $\text{P}(Mexican|A) = \frac{3+1}{4+2} = \frac{4}{6} = \frac{2}{3}$
    - Cooking:
        - $\text{P}(Grilled|A) = \frac{2+1}{4+3} = \frac{3}{7}$
        - $\text{P}(Steamed|A) = \frac{0+1}{4+3} = \frac{1}{7}$

- $P(Fried|A) = \frac{2+1}{4+3} = \frac{3}{7}$

o Spiced:

- $P(Spicy|A) = \frac{2+1}{4+2} = \frac{3}{6} = \frac{1}{2}$
- $P(Normal|A) = \frac{2+1}{4+2} = \frac{3}{6} = \frac{1}{2}$

• עבור A' (Laplance):

o Theme:

- $P(Indian|A') = \frac{4+1}{6+2} = \frac{5}{8}$
- $P(Mexican|A') = \frac{2+1}{6+2} = \frac{3}{8}$

o Cooking:

- $P(Grilled|A') = \frac{3+1}{6+3} = \frac{4}{9}$
- $P(Steamed|A') = \frac{1+1}{6+3} = \frac{2}{9}$
- $P(Fried|A') = \frac{2+1}{6+3} = \frac{3}{9} = \frac{1}{3}$

o Spiced:

- $P(Spicy|A') = \frac{3+1}{6+2} = \frac{4}{8} = \frac{1}{2}$
- $P(Normal|A') = \frac{3+1}{6+2} = \frac{4}{8} = \frac{1}{2}$

a) נחשב עבור Normal, Grilled, Mexican:

- $P(A|Normal, Grilled, Mexican) = P(A) * P(Normal|A) * P(Grilled|A) * (P(Mexican|A) = \frac{2}{5} * \frac{1}{2} * \frac{3}{7} * \frac{2}{3} = \frac{12}{210} = \frac{2}{35}$
- $P(A'|Normal, Grilled, Mexican) = P(A') * P(Normal|A') * P(Grilled|A') * (P(Mexican|A') = \frac{3}{5} * \frac{1}{2} * \frac{4}{9} * \frac{3}{8} = \frac{36}{720} = \frac{1}{20}$

ננרמל:

- $P(A|Normal, Grilled, Mexican) = \frac{\frac{2}{35}}{\frac{1}{20}+\frac{2}{35}} = 0.5333$
- $P(A'|Normal, Grilled, Mexican) = \frac{\frac{1}{20}}{\frac{1}{20}+\frac{2}{35}} = 0.4666$
- $P(A|Normal, Grilled, Mexican) = 0.5333 > 0.4666 = P(A'|Normal, Grilled, Mexican)$

ולכן נצפה שהמנה תוחזר ולכן נמליץ לבעל המסעדה לא להגיש אותה.

b) נחשב עבור Spicy, Steamed, Indian:

- $\mathrm{P}(A|Spicy, Steamed, Indian) = P(A) * P(Spicy|A) * P(Steamed|A) * (P(Indian|A) = \frac{2}{5} * \frac{1}{2} * \frac{1}{3} * \frac{1}{7} = \frac{2}{210} = \frac{1}{105}$
- $\mathrm{P}(A'|Spicy, Steamed, Indian) = P(A') * P(Spicy|A') * P(Steamed|A') * (P(Indian|A') = \frac{3}{5} * \frac{1}{2} * \frac{2}{9} * \frac{5}{8} = \frac{30}{720} = \frac{1}{24}$

ננרמל:

- $\mathrm{P}(A|Spicy, Steamed, Indian) = \frac{\frac{1}{105}}{\frac{1}{105}+\frac{1}{24}} = 0.1860$
- $\mathrm{P}(A'|Spicy, Steamed, Indian) = \frac{\frac{1}{24}}{\frac{1}{105}+\frac{1}{24}} = 0.8139$
- $\mathrm{P}(A|Spicy, Steamed, Indian) = 0.1860 < 0.8139 = \mathrm{P}(A'|Spicy, Steamed, Indian)$

ולכן נצפה שהמנה לא תוחזר ולכן נמליץ לבעל המסעדה כן להגיש אותה.