# תרגיל בית שלישי- מבוא ללמידה ממוכנת

**1. פתרון:**

**א.** מעגל $(r,c)$ מוגדר ע"י מרכזו, $c$, וע"י הרדיוס שלו, $r$. נסתכל על משפחת המסווגים ההבאה:

$$H=\{h_{r,c}:r\in\mathbb{R},c\in\mathbb{R}^2\}$$

$$h_{r,c}(x)=\begin{cases}1 & \|x-c\|_2\le r\\ 0 & else\end{cases}$$

נמצא את ה$VCdim$ של המשפחה עם הוכחה מלאה.

נראה $VCdim(H)=3$, כלומר נראה שקיימות 3 נקודות שניתנות לניפוץ, ולאחר מכן נראה כי אין 4 נקודות שניתנות לניפוץ.

דוגמה ל3 נקודות שניתנות לניפוץ:

$$a=(0,0),b=(0,2),c=(2,0)$$

| $h_{r,c}(a)$ | $h_{r,c}(b)$ | $h_{r,c}(c)$ | $r$ | $c$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0.5 | $(-1,-1)$ |
| 0 | 0 | 1 | 1 | $(0,0)$ |
| 0 | 1 | 0 | 1 | $(0,2)$ |
| 0 | 1 | 1 | 2 | $(2,2)$ |
| 1 | 0 | 0 | 1 | $(2,0)$ |
| 1 | 0 | 1 | 1 | $(0,1)$ |
| 1 | 1 | 0 | 1 | $(0,2)$ |
| 1 | 1 | 1 | 2 | $(1,1)$ |

ועכשיו נראה שאין 4 נקודות שניתנות לניפוץ:

ניקח 4 נקודות שרירותיות, $a,b,c,d$. נניח בלי הגבלת הכלליות שניתן להקיף כל נקודה, כל 3 נקודות ואת כל הנקודות, ונניח בשלילה שגם ניתן להקיף כל זוג נקודות ונראה שמתקיימת סתירה.

ישנם 6 זוגות של נקודות 4 מהם חייבים להתקיים כיוון שהנקודות אינן נמצאות על קו ישר. ועכשיו נשארו 2 זוגות שהם למעשה יוצרים את האלכסונים של המרובע שנוצר ע"י 4 הנקודות. נניח שאחד מהזוגות ניתן להקיף ע"י מעגל (בלי ש2 הנקודות הנותרות יהיו בתוכו), כלומר אורך האלכסון שלו קטן מהזוג השני. כלומר כאשר ננסה להקיף את הזוג השני בהכרח אחת מהנקודות של הזוג הראשון ימצאו במעגל כי האלכסון של הזוג השני יותר גדול מהלכסון הראשון. ולכן יש פה סתירה בניגוד להנחה שלנו ולכן מתקיים $VCdim(H)\le 3$.

והראנו מקודם שמתקיים $VCdim(H)\ge 3$ ולכן מתקיים $VCdim(H)=3$.

**ב.** מהי לכל היותר כמות הדגימות שאנו צריכים עבור $H$ש תהיה $agnostic\ PAC\ learnable$, כך שמתקיים:

$$\delta_H = e^{-2}, \varepsilon_H = \frac{\varepsilon}{2}$$

נשתמש בנוסחה הבאה:

$$m \geq \left\lceil \frac{2\log\left(\frac{2|H|}{\delta}\right)}{\varepsilon^2} \right\rceil$$

וגם נזכור שמתקיים $|H| \leq 2^{VCdim(H)}$ לכן מתקיים הדבר הבא:

$$m \geq \left\lceil \frac{2\log\left(\frac{2\cdot 2^{VCdim(H)}}{\delta_H}\right)}{\varepsilon_H^2} \right\rceil = \left\lceil \frac{2\log\left(\frac{2\cdot 2^3}{e^{-2}}\right)}{\left(\frac{\varepsilon}{2}\right)^2} \right\rceil = \left\lceil \frac{2\,(\log(8)+2)}{\frac{\varepsilon^2}{4}} \right\rceil = \left\lceil \frac{8\,(\log(8)+2)}{\varepsilon^2} \right\rceil$$

$$= \left\lceil \frac{32.636}{\varepsilon^2} \right\rceil$$

**ג.** פתרון:

$$m \geq C_1 \cdot \frac{1}{\epsilon}\left(VCdim(H) + \ln\frac{1}{\delta}\right)$$

$$m \geq C_1 \cdot \frac{1}{\frac{\epsilon}{2}}\left(3 + \ln\frac{1}{e^{-2}}\right) = \frac{2}{\epsilon}(3+2) = \frac{10}{\epsilon}$$

לכן, צריך לפחות $\frac{10}{\epsilon}$ דגימות.

**2. פתרון:**

אנחנו רוצים לקנות מכונית.

מידע המכוניות מחולק לכמה מאפיינים: $buying$ – מחיר הקנייה, $maint$ – מחיר התחזוקה, $doors$ – דלתות, $persons$ – המספר המקסימלי של האנשים שהמכונית יכולה להכיל, $lug\_boot$ – גודל המטען, $safety$ – מאובטח.

כל מכונית ממופת לאחר מארבעת הקטגוריות: $unacc, acc, good, vgood$.

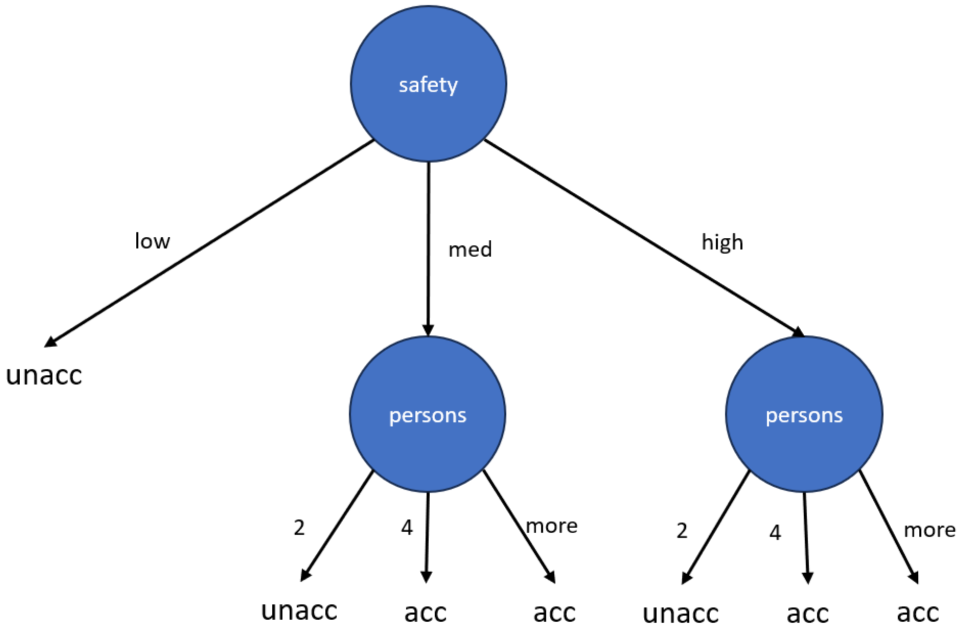
**א.** בקוד

**ב.** הדיוק של ה$decision\ tree$ על מידע האימון הוא $100\%$. כשלעץ אין דיוק מושלם זה כיוון שיש לו 2 אובייקטים בעלי אותו תכונות שאנחנו בודקים אבל לא בעלי אותו תוויות. עבור המידע שלנו אין 2 מכוניות בעלי אותו תכונות שאנחנו בודקים ואין להם את אותו תווית . ולכן הדיוק של העץ הוא מושלם.

**ג.** כאשר אנחנו מגבילים את עומק לפחות מ6 יש מכוניות בעלי אותו תכונות שאנחנו בחרנו לא בעלי אותו תוויות ולכן לא באמת יכול להיווצר עץ מושלם. וכמובן שאנחנו מריצים העץ מתאמן על מידע האימון שלנו והוא ינסה

כמה שיותר תואם עובר מידע האימון שלנו, לכן לרוב הדיוק של העץ על מידע האימון הוא גדול מהדיוק על מידע הבדיקה כמו שיצא לנו בתוצאות מהקוד.

העץ שיוצא עבור עץ בעומק 2 עבור כל המידע כאשר דיוק העץ הוא 77.778%, הוא:




**3. פתרון:**

**א.** נניח שיש קבוצה של משתנים מקריים בלתי תלויים, $\{Y_i\}_{i=1}^{n}$, עם התוחלת $\mu$ ועם סטיית התקן $\sigma^2$. כאשר לכל משתנה מקרי, $Y_i$ , יש מסווג $i$.

עכשיו נראה את התוחלת וסטיית התקן של הממוצע של המשתנים המקריים, $Y = \frac{1}{n}\sum_{i=1}^{n} Y_i$, ונראה למה התשובה שלנו מרמזת שכדאי לעבוד עם מסווג שהוא הממוצע של המסווגים מאשר לעבוד עם מסווג $i$ ספציפי.

קודם כל נמצא את התוחלת ואת סטיית התקן של $Y$:

$$E[Y] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$Var(Y) = Var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) =_1 \frac{1}{n^2}Var\left(\sum_{i=1}^{n} Y_i\right) =_2 \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

1) $Var(aX) = a^2 Var(X)$

2) אם $X, Y$ בלתי תלויים אזי $Var(X+Y) = Var(X) + Var(Y)$.

עכשיו ניתן לראות שההבדל בין משתנה מקרי $Y_i$ לבין $Y$ הוא סטיית התקן שלהם, כאשר ל$Y_i$ יש סטיית תקן יותר גדולה מאשר $Y_i$. לכן כאשר ננסה לסווג לפי המסווג שהוא המסווג של הממוצעים סיכוי ההצלחה שלו גבוה יותר מאשר מסווג ספציפי.

**ב.**

- המודל שנותן שגיאה יותר גדולה על מידע האימון הוא ה$Random\ Forest$.
- השגיאה היא לא מאוד משמעותית בין 2 המודלים אבל כן קיימת שגיאה ברורה בין 2 המודלים.
- ב$Random\ Forest$ ההכללה על מידע האימון היא יותר גדולה ולכן גם השגיאה על המידע האימון תהיה יותר גדולה. אבל גם יהיה פחות $overfitting$ על המידע שלנו ולכן גם השגיאה תהיה יותר קטנה על מידע הבדיקה שלנו.

**ג.** לפי הגרף מספר ה$estimators$ הכי טוב הוא בין 1 ל4.

**4. פתרון:**

**א.** ראשית, נמזער את פונקציית המטרה $J_{SSE}$. נגזור לפי $\mu_i$ ונשווה ל0:

$$\frac{dJ_{SSE}}{d\mu_i} = -2\sum_{x\in D_i}(x-\mu_i)$$

$$-2\sum_{x\in D_i}(x-\mu_i) = 0$$

$$\sum_{x\in D_i} x - \sum_{x\in D_i}\mu_i = 0$$

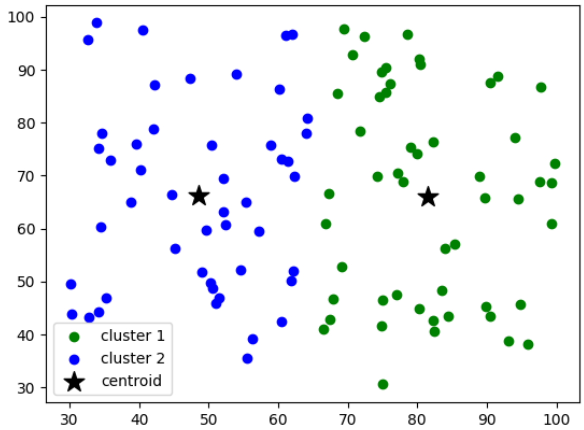
$$\frac{1}{n_i}\cdot\sum_{x\in D_i} x = \frac{1}{n_i}\cdot\sum_{x\in D_i}\mu_i$$

ולכן:

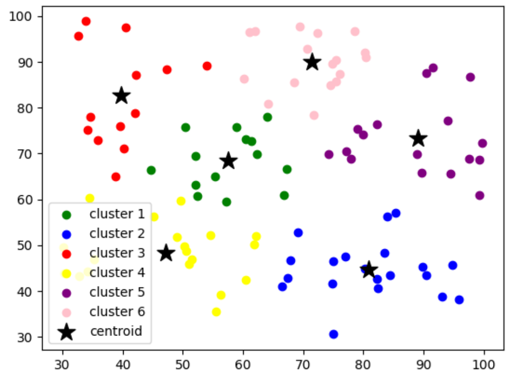
$$\boldsymbol{\mu_i = \frac{1}{n_i}\cdot\sum_{x\in D_i} x}$$

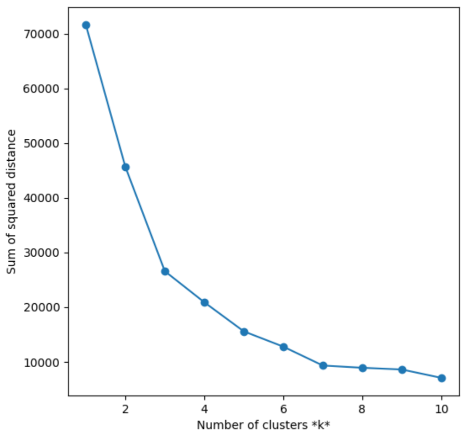
כנדרש.

**ב.**




**ג.** בחרנו $k = 6$ כי כפי שניתן לראות בגרף משמאל, החל מ$k = 6$, שיפוע הגרף מתחיל להיות מתון.




**ד.**