

שאלה 1:

בסט האימון המסווג למד את אופי התוצאה המבוקשת לעומת סט הבדיקה ולכן תוצאות יהיו מדויקות יותר

Train_table

1	1000	30	500	500	1	1	100	100	Max_depth
100	1000	400	1	200	1	100	1	100	N_estimators
0.47	1	1	0.95	1	0.43	0.47	0.95	1	res

Check_table

1	1000	30	500	500	1	1	100	100	Max_depth
100	1000	400	1	200	1	100	1	100	N_estimators
0.45	0.95	0.95	0.86	0.95	0.45	0.45	0.86	0.95	res

1. מסקנה: ככל שעומק העץ גבוה יותר ומספר העצים גבוה יותר התוצאה תהיה מדויקת יותר (והגורם המשפיע ביותר הוא עומק העץ)

2. התוצאות מדויקות יותר ב train_table לעומת ה check_table אבל בשניהם היחס נשמר אותו דבר וככל שמספר העצים קטן יותר הפער גדל

כאשר תוצאות הסיווג נמוכות

```
array([[0.55966209, 0.26399155, 0.12988384, 0.04646251],
       [0.08418367, 0.22048105, 0.38447522, 0.31086006],
       [0.55966209, 0.26399155, 0.12988384, 0.04646251],
       ...,
       [0.08418367, 0.22048105, 0.38447522, 0.31086006],
       [0.08418367, 0.22048105, 0.38447522, 0.31086006],
       [0.55966209, 0.26399155, 0.12988384, 0.04646251]])
```

```
array([[0.01081731, 0.20432692, 0.421875, 0.36298077],
       [0.42837929, 0.24983188, 0.20477471, 0.11701412],
       [0.42837929, 0.24983188, 0.20477471, 0.11701412],
       ...,
       [0.01081731, 0.20432692, 0.421875, 0.36298077],
       [0.42837929, 0.24983188, 0.20477471, 0.11701412],
       [0.42837929, 0.24983188, 0.20477471, 0.11701412]])
```

כאשר תוצאות הסיווג גבוהות

```
array([[0.99, 0.01, 0., 0. ],
       [0., 0., 0.99, 0.01],
       [1., 0., 0., 0. ],
       ...,
       [0., 0., 0.02, 0.98],
       [0., 0., 0.96, 0.04],
       [1., 0., 0., 0. ]])
```

train

```
array([[0.01, 0.8, 0.18, 0.01],
       [0., 0.02, 0.97, 0.01],
       [0., 0.15, 0.83, 0.02],
       ...,
       [0., 1., 0., 0. ],
       [1., 0., 0., 0. ],
       [0., 0.1, 0.46, 0.44]])
```

check

מסקנה: ככל תוצאות הסיווג גבוהות יותר ההתפלגויות ממוקדות יותר

שאלה 2:

1. רמת הדיוק היא 0.621
2. הפרמטרים החשובים: $velX_9, posZ_0, posZ_5, velZ_7, velX_8, posZ_6$
3. Check_table:

				100	50	10	5	3	Max_depth
				0.87	0.88	0.89	0.755	0.621	res

Train_table:

				100	50	10	5	3	Max_depth
				1	1	0.95	0.789	0.63	res

ההבדל גדול יותר ב - decision tree

שאלה 3:

1. דיוק על סט האימון: 0.319
דיוק על סט הבדיקה: 0.286
ההבדל ביניהם נובע מ overfitting
- 2.

100		40		12		1		C
סט בדיקה	סט אימון	סט בדיקה	סט אימון	סט בדיקה	סט אימון	סט בדיקה	סט אימון	score
0.289	0.318	0.287	0.320	0.294	0.322	0.286	0.319	

3. המסווג מוצא 4 משקולות