

עקרונות מדעי הנתונים

תרגיל מס' 3

תאריך ההגשה (דרך המודל): 18.12.2021

שאלה 1 (33 נק')

חיים אופה לחם ומעוניין להבין מה כמות השמרים האופטימלית בגרמים שעליו לשים בבצק שגודלו קבוע על מנת להגיע ללחם מוצלח, תופח ואוורירי. חיים החליט לבצע מספר ניסויים בהם המשתנה היחיד שערכיו השתנו היה כאמור כמות השמרים בבצק (Amount of yeast).

להלן הטבלה הבאה המציגה 5 ניסויים שערך חיים, כמות הגרמים של השמרים בבצק הקבוע וה-class שאומר:

0 – בצק לא מוצלח

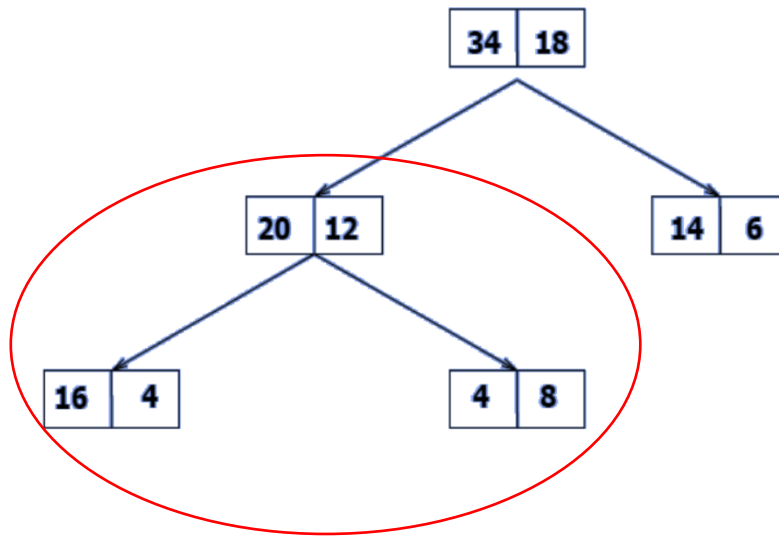
1 – בצק מוצלח ואוורירי.

Record	Yeast (gr)	Class
1	1.5	1
2	1	0
3	1.7	1
4	1.5	0
5	2.1	1

חשבו את ה-Threshold האופטימלי באמצעות Discretization Algorithm שנלמד בהרצאה (בעית הטמפרטורה וה-Decision בין Shopping, Cinema ו-Tennis).

שאלה 2 (33 נק')

חיים בנה עץ החלטה באמצעות אלגוריתם כלשהו. להלן העץ שנבנה.



לאחר מכן החליט לנסות לשפרו ולנסות למנוע ממנו לסבול מ-overfitting. לשם כך, החליט להפעיל עליו Pessimistic Error Pruning על הענף המוצג בעיגול. האם על פי האלגוריתם חיים יקצוץ את הענף או שלא?

חשבו את האלגוריתם בעצמכם והציגו זאת בעבודה.

שאלה 3 (33 נק')

נתונה הטבלה הבאה המציגה נתוני קניות הסטוריים לגבי רכישת בתים:

ID	Income	Season	District	House Type	Num of floors	Bought?
1	High	Autumn	Suburban	Detached	1	No
2	High	Summer	Suburban	Detached	2	No
3	High	Spring	Rural	Detached	1	Yes
4	High	Winter	Urban	Semi-detached	1	Yes
5	Low	Spring	Urban	Semi-detached	1	Yes
6	Low	Autumn	Urban	Semi-detached	2	No
7	Low	Autumn	Rural	Semi-detached	2	Yes
8	High	Spring	Suburban	Terrace	1	No
9	High	Summer	Urban	Terrace	1	No
10	Low	Winter	Urban	Terrace	1	Yes
11	Low	Spring	Suburban	Semi-detached	1	Yes
12	Low	Autumn	Suburban	Terrace	2	Yes
13	High	Autumn	Rural	Terrace	2	Yes
14	Low	Spring	Rural	Detached	1	Yes
15	High	Spring	Urban	Terrace	2	No

השתמשו באלגוריתם Naïve Bayes ובנתונים שבטבלה העליונה כדי לקבל החלטה למי מהלקוחות החדשים כדאי לשלוח פרסומת על מנת שיקנה דירה:

ID	Income	Season	District	House Type	Num of floors	Bought?
16	High	Autumn	Suburban	Detached	2	?
17	High	Spring	Rural	Detached	2	?
18	Low	Spring	Rural	Semi-detached	1	?

יש להשתמש ב-Laplacian estimate.

הוראות הגשה:

1. עליכם להגיש קובץ Word הכולל את מספרי תעודות הזהות של שני המגישים עם קו תחתון מפריד ביניהם בצורה הבאה: ID1_ID2.docx.
2. את הקובץ docx שייצרתם עליכם להעלות לתיבת ההגשה.
3. הגשה **בזוגות בלבד**.

בהצלחה!!