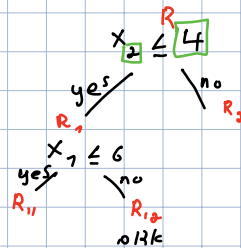
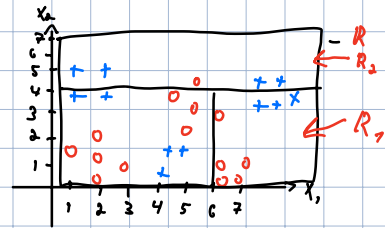
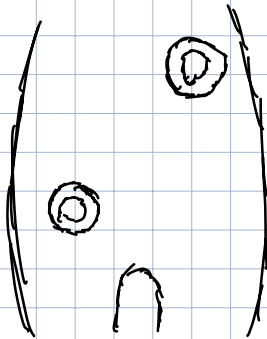


38. כתוב: (ID3)

דוגמה - הבה if else על שני דוגמאות

תחילת



(1) אם האלגוריתם - ביצור התחביר לפני חלקים  $R_1, R_2$  דוגמת "מאפי פסל" -  $x_j \leq t$

j-מחלק על כל התבונה  
t-מחלק על כל הדגים  
שמופיעים בעמדה j

$$S(j, t) = (R_1, R_2)$$

$$\{x \mid x_j \leq t, x \in R\} \quad \{x \mid x_j > t, x \in R\}$$

(2) בעצם קוצים על בינארי - הדגים הם הכחלול (שם של מחלקה) ושני התבונות הם הדגים של הערצות

(3) המחלקה - קוצים בעצרת הפיצולים לפימרים עם מחלקה יחידה ונרצה פונ' מחיר להתחבר על הכיוון הנכון.

נרצה מפתחיות המחיר  $C(R)$  שיתקיים:

ערור, נגיד ב מחלקות אם הם הדגמאות ב-R מחלקות 50%, 50%, 50% המחבר "איד"  $C(R)$  (ידור) ואם הם הדגמאות מחלקה 100%, 0% מחבר "אשוק"  $C(R)$  (מחלק).

2 פונקציות  $C(R)$  אופטימליות:

(1) פונקציית האינפורמציית

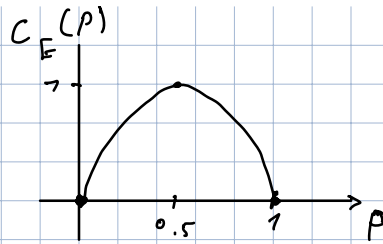
$$C_E(R) = - \sum_{i \text{ מחלקה}} p_i \log_2 p_i$$

בערוב:  $0 \cdot \log_2 0 = 0$   
 $\lim_{p \rightarrow 0} p \log_2 p = 0$

$$(p, 1-p) \text{ מחלקה דגם מחלקה}$$

$$-(p \log_2 p + (1-p) \log_2 (1-p))$$

$$p=0.5: -(0.5 \log_2 0.5 + 0.5 \log_2 0.5)$$

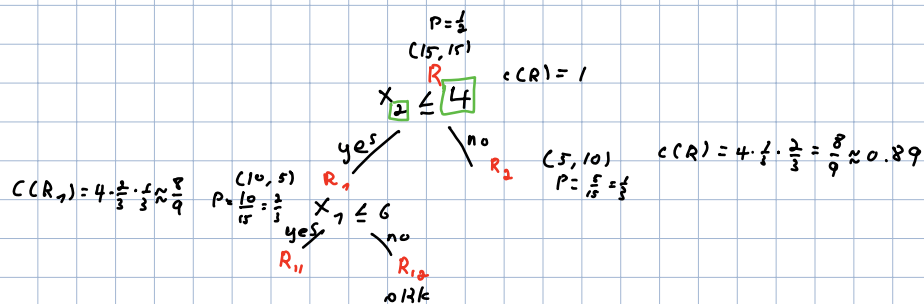
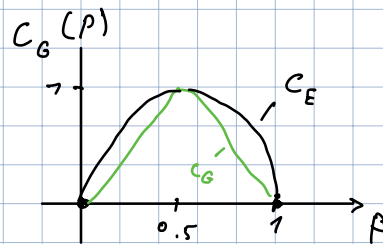


Gini  $\approx 32\%$  (2)

$$C_G(R) = 2 \sum_{i=1}^n p_i (1 - p_i) \quad \left[ = 2 \left( \sum_{i=1}^n p_i - \sum_{i=1}^n p_i^2 \right) = 2 \left( 1 - \sum_{i=1}^n p_i^2 \right) \right]$$

$p_1 = p, p_2 = 1 - p$  : שני תוצאות

$$2(p(1-p) + (1-p)p) = 4p(1-p) = F(p)$$



נרצח עמדת בכתב שפרט את כל המצבים / מצב נרצח  
 $|R| \cdot \frac{1}{|R|} C(R_1) + |R_2| \cdot \frac{1}{|R|} C(R_2)$   
 פה רואים שהכלל הזה הוא זהה לזה של ה- $C$  אבל עם משקל.  
 אז אפשר להגדיר את  $C$  כממוצע ממשוקלל של  $C$  על המצבים הנרצחים.

$$1 - \left( \frac{15 \cdot 0.89 + 15 \cdot 0.89}{30} \right) = 0.11 \text{ מיליון}$$

$$C(R_{1,2}) = 0$$

$$C(R_{1,1}) = 1$$

$$:R_1 \rightarrow \text{?/c?}$$

$$0.89 - \left( \frac{10 \cdot 1 + 5 \cdot 0}{15} \right) = 0.89 - \frac{10}{15} \approx 0.23$$

מדקן פֿון כתובים/התלוקות  $S(j, t)$  נדמר בהתלוקה שניתן הקסימאט נספיר, וכן, הפאק דרקריס'ג.

### תוספות/כעדות

(1) זמן חישוב - פאקא היז' יס פנו מ.מ. אסמאט. כך למל ויז' על-סג פל ה-  $R_i$  פל דמאר דס פל האלפא ולכן מ.מ. אפסיוות תלוקה (דיוק).

ולכן סג זמן חישוב

דיוק - מ.מ. מ.

$$d = \log m, \frac{m}{d}$$

(2) הנית על הולא תפסיה רשסיה -

סדור פלפא דס כח פאפאט, נחא אה פסרי המאפד.

$$\bar{y} = \frac{\sum_{i \in R} y_i}{|R|}$$

$$C(R) = \frac{\sum_{i \in R} (y_i - \bar{y})^2}{|R|}$$

$C(R)$  דלוק אה  
פסוואט גלוקה  
 $C(R)$  נאק  
אה פסוואט  
נאפד

(3) פאפאט: אה זמן חישוב פל  $\sum_{i \in R} y_i$

(א) פל פספיר אה המאפד  
(ב) פלפא גלוקה תלוקה זל פאפאט למנו מספיר אה קאלי.

אפסיה  
פספיר קאלי

מסלול: Overfitting -  
פסוואט פסוואט > trainset  
Variance דלוק

(4) פאפאט דקאפסיה Overfitting:

(א) פאפאט פלפא פסוואט - למל לא נפול דקאפסיה דס פאפאט מ-סב נהקרי  
(ב) פאפאט זל מוספלי  
(ג) פסס פסל מל דקאפסיה דלוק.

(5) Random Forest - נאק 100 זצי דהולא (פאפאט...), וליק אה פאפאט דקאפסיה

פאפאט נהקרי דקאפסיה

(א) פל פסס נהקרי זל מוספלי

(ב) דלוק  $S(j, t)$  נחא פל מוספלי מוספלי



L.

