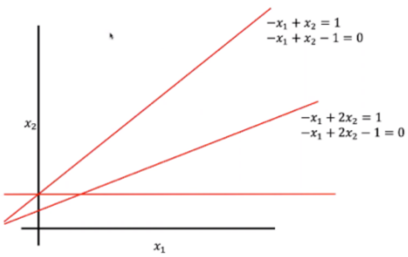
למידה חישובית ממידע: תרגול 2 – Decision Trees

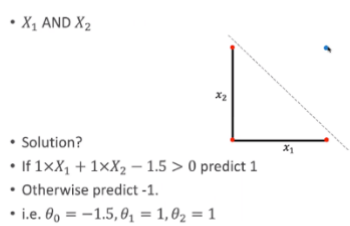
חזרה קצרה על מה שנלמד בתרגול הקודם:

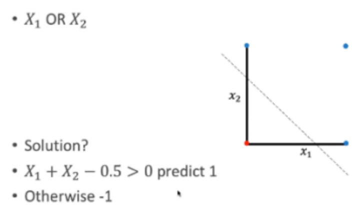
- גרדיאנט - וקטור הנגזרות החלקיות
- עבור פונקציה רבת משתנים $f(x_1, x_2, \dots, x_n)$ הגרדיאנט עבורה יהיה: $\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)$
- הגרדיאנט מצביע לכיוון העלייה התלולה/הגדולה ביותר ביחס לנקודה בה הוא מחושב, וערכו הוא השיפוע בכיוון זה. כך שאם נלך נגד כיוון הגרדיאנט, נתקדם לכיוון המינימום (גרדיאנט דיסנט).
- בנוסף דיברנו על הצורך ב- learning rate, אלפא, במטרה לקבוע כמה מהר או לאט נתקדם לכיוון המינימום (משקל אופטימלי). העלנו את הדיון הבא: מצד אחד לא נרצה אלפא גדולה מידי כדי שלא נגיע למצב בו "דילגנו" על פני המינימום, וכן מצד שני, לא נרצה אלפא קטנה מידי כדי להגיע למינימום בזמן סביר.

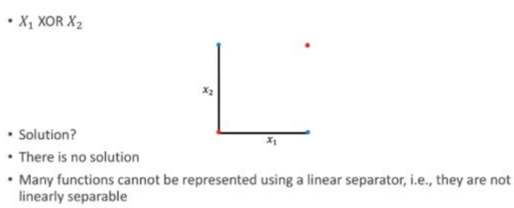
תזכורת באלגברה:



- **מפריד לינארי** הוא למעשה היפר-מישור במרחב שמוגדר על ידי הווקטור טטה, וכל הנקודות על ההיפר-מישור הנ"ל פותרות את המשוואה $\theta_1 x_1 + \dots + \theta_n x_n = b \quad (= \theta_0)$ כאשר x הן הקואורדינטות של הנקודה.
- ההיפר-מישור מפריד את המרחב לשני חללים, כל הנקודות שהמשוואה פותרת הן מעל b.
- הטטות שולטות בזווית של המישור.
- הטטה0 מזיזה מראשית הצירים.
- נרצה למצוא מפריד לינארי: כל נקודה עליונה עם תוצאה שגדולה מ-0, תהיה שייכת למחלקה 1+ או (1-). כל נקודה תחתונה עם תוצאה נמוכה מ-0, תהיה שייכת למחלקה 1- (או 1+).
- לכן עלינו למצוא את הווקטור טטה (שיש בו n+1 ערכים, n משקלים וביאס טטה0).
- נחזה 1 אם $\sum_{i=1}^{n} \theta_i x_i + \theta_0 > 0$ ואחרת 1-.
- **להלן דוגמאות:**



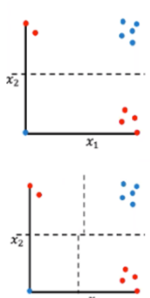




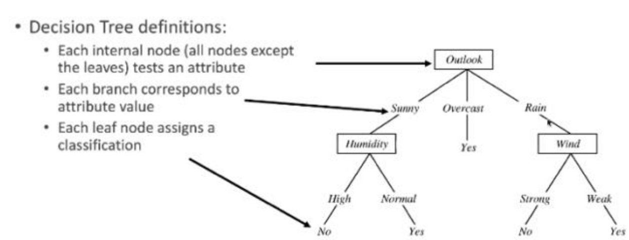
העניין הוא שלא כל דאטה ניתן להפרדה לינארית (להלן בעיית הקסור)
ולכן נצטרך כלים קצת יותר חזקים ממפרידים לינאריים (נרחיב עליהם בהמשך).

עצי החלטה

האינטואיציה של עצי החלטה: האטריביוטים שיבחרו, יבחרו על פי מי שמקרב אותנו ביותר להפרדה מלאה בילדים. נרצה למצוא חלוקה לבנים, כך שלאחר החלוקה נהיה כמה שיותר קרובים לטהורות בבנים.



- עלה הוא תשובה לקלסיפיקציה
- כל ענף מייצג ערך של האטריביוט הנבחר
- כל צומת פנימי בוחנים אטריביוט.
- יודע לטפל גם באטריביוטים רציפים, עבורם נבחר threshold במטרה לחלק לאטריביוטים.



האלגוריתם לבניית העץ: (דוגמה לעץ שנבנה בתור, יש אפשרות לבנות עץ גם ברקורסיה)

- כל עוד יש צמתים בתור תבצע:
  -- קח את הצומת הבא בתור.
  -- אם הדגימות בצומת זה טהורים (שייכים כולם לאותו הקלאס) המשך לצומת הבא
  -- אחרת: נכניס ל-A את האטריביוט הטוב ביותר עבור קבוצת הדגימות בצומת n, נמנה את A לאטריביוט ההחלטה של הצומת n, ולכל ערך של האטריביוט A ניצור בן חדש לצומת n. נפלג את הדגימות של צומת n לבנים החדשים שלו ונכניס את הצמתים הבנים לתור.
- כשאין יותר בנים בתור נסיים את הלולאה

איך בוחרים את **האטריביוט הטוב ביותר** שיקרב אותנו כמה שיותר לטהורות/וודאות? כדי לענות על שאלה זו ראשית עלינו לבחור מדד שיגיד לנו האם הפיצול הוא אכן טוב. עלינו למדוד כמה אנחנו רחוקים מטהורות (perfect classification), **מדידה זו נקראת impurity (מחושב על צומת):**

- Impurity/ אי טהורות גבוהה משמעותה – אנחנו רחוקים מקלסיפיקציה מושלמת
  (ככל שיותר קרובים להתפלגות יוניפורמית נתרחק יותר מקלסיפיקציה מושלמת)
- Impurity/ אי טהורות נמוכה משמעותה – אנחנו קרובים לקלסיפיקציה מושלמת (בהרצאה מופיעה הגדרה פורמלית יותר)

<u>בחירת האטריביוט הטוב ביותר</u>

- נחשב את רמת הטהורות Impurity עבור הצומת הנוכחי (בו אנו נמצאים)
- נחשב את הממוצע המשוקלל של רמת הטהורות על הצמתים-הבנים לאחר פיצול לפי האטריביוט שנבדק
- נחסיר את השני(הממוצע המשוקלל) מהראשון(רמת הטהורות של האבא) ונקבל את הפרש ה-impurity
- נבחר את האטריביוט שמניב את הפרש ה-Impurity המקסימלי, הגבוהה ביותר
- **הפרש ה-Impurity בין רמת הטהורות של צומת האב לבין ממוצע רמת הטהורות שמתקבלת עבור הבנים לאחר הפיצול נקראת Goodness of split.**

הנוסחה שמחשבת את הפרש ה-Impurit מקבלת שני אינפוטים; את האטריביוט הנבדק A ותת קבוצת הדגימות אשר בצומת הנוכחי:

$$\Delta\varphi(S,A) = \varphi(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|}\varphi(S_v)$$

* Where φ is the impurity measure

רמת הטהורות בצומת האב פחות הממוצע הממושקל של הטהורות בבנים הנוצרים מהפיצול על פי A (יש לשים לב שהאימפיוריטי מחושב על הקלאס ואין שום קשר לערכי האטריביוט כשזה מגיע לחישוב האימפיוריטי מלבד הפיצול לבנים על פי האטריביוט).

<u>עובדות חשובות: (האות היוונית פי מסמלת את ה-impurity)</u>

- מדד אי הטהורות (impurity measure) מודד לפי התפלגות הקלאסים בצומת כלשהו
- הדגימות מפוצלות לפי ערכי האטריביוט הנבדק – A
- זאת אומרת שאנו מפצלים את הדגימות לפי ערכי האטריביוט ואז מחשבים את רמת הטהורות לפי ערכי הקלאסים.

<u>טיב הפיצול / Goodness of Split</u>

ישנם 2 יישומים עיקריים עבור קריטריון הטהורות:

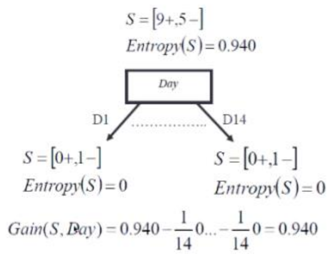
	Gini	Entropy
Impurity	$GiniIndex(S) = 1 - \sum_{i=1}^{c}\left(\frac{\lvert S_i\rvert}{\lvert S\rvert}\right)^2$	$Entropy(S) = -\sum_{i=1}^{c}\frac{\lvert S_i\rvert}{\lvert S\rvert}log\frac{\lvert S_i\rvert}{\lvert S\rvert}$
Goodness of split	$Gini\ Gain = GiniIndex(S) - \sum_{v \in Values(A)}\frac{\lvert S_v\rvert}{\lvert S\rvert}GiniIndex(S_v)$	$Information\ Gain = Entropy(S) - \sum_{v \in Values(A)}\frac{\lvert S_v\rvert}{\lvert S\rvert}Entropy(S_v)$

- C מסמל את מספר הקלאסים
- הערך המקסימלי שג'יני אינדקס יכול לקבל הוא 1, ככל שכמות הקלאסים עולה הסכום שואף ל-0. כאשר יש שני קלאסים המקסימום שג'יני יקבל יהיה 0.5
- הערך המקסימלי של אנטרופיה עבור 2 קלאסים הוא 1, עבור 4 קלאסים הוא 2. מה שניתן להבין מכאן הוא שהאנטרופי לא חסום.
- ערכם המינימלי הוא 0.
- את שתי הנוסחאות נמיר לפורמולה של גודנס אוף ספליט ובהן נשתמש (שורה שניה בטבלה).

התמודדות עם אטריביוט עם ערכים רבים

• Imagine using the attribute DAY=[D1,...,.D14]

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



- בהינתן דאטה של ציונים בקורס, יש לנו 100 דגימות של סטודנטים עם 5 אטריביוטים: תעודת זהות, מין, ממוצע בגרות, שעות למידה, מזל (1-10), נרצה לחזות מי יעבור את המבחן ומי לא.
- באיזה אטריביוט יבחר InformationGain ולמה? נשים לב שיבחר האטריביוט של תעודת הזהות מפני שהוא בעל 100 ערכים שונים, הוא יעניק לך תשובה לכל סטודנט. יש לו תשובה ייחודית והוא יוריד את האי וודאות ב-100%. אבל כמובן שזו בחירה לא טובה, ובאופן כללי נרצה להכניס דאטה חדש ולחזות עליו, מה שלא יתאפשר עבור ת״ז שלא נמצא במאגר שלי.
- יש נטייה באלגוריתם של גודנס אוף ספליט לבחור באטריביוט בעל ערכים רבים.
- **הדרך להימנע מנטייה זו היא להשתמש ב- GainRatio:**

$$GainRatio(S,A) = \frac{InformationGain(S,A)}{SplitInformation(S,A)}$$

ניקח את ה-InformationGain(S,A) ו״ננרמל״ אותו על ידי חילוק ב-SplitInformation(S,A).

$$SplitInformation(S,A) = -\sum_{a\in A}\frac{|S_a|}{|S|}log\frac{|S_a|}{|S|}$$

כאשר SplitInformation(S,A) הוא האנטרופיה ביחס לאטריביוט A:

עד כה חישבנו אנטרופיה **עבור קלאס**, כעת נחשב אנטרופיה **בהתחשב באטריביוט**. ככל שיהיו לנו יותר ערכים באטריביוט ככה נגיע למקרה שהוא יותר יוניפורמי. (ככה ״מענישים״ אטריביוט בעל ערכים רבים)

$$SplitInformation(S,Day) = -\sum_{i=1}^{14}\frac{1}{14}log\frac{1}{14} = -log\frac{1}{14} = 3.8074$$

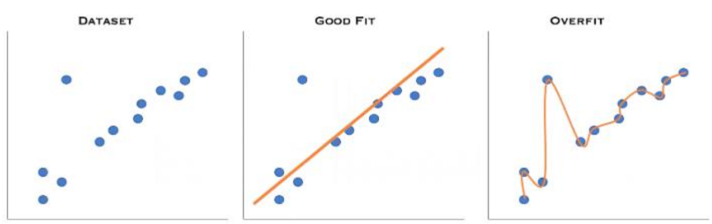
$$GainRatio(S,A) = \frac{0.94}{3.8074} = 0.2469$$

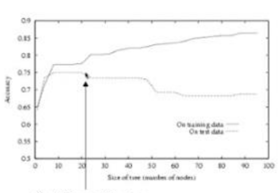
דוגמה לחישוב ספליט אינפורמיישן וגאין רשיו עבור הדוגמה המוצגת לעיל בתמונה, ועבור האטריביוט Day.

אוברפיטינג / Overfitting

בשלב מסוים באלגוריתם הלמידה ככל שהופכים את המודל שלנו למורכב יותר, כך אנחנו מתאימים יותר את המודל שלנו לדאטה שלנו. אנחנו נרצה להגיע למודל כמה שיותר כללי כך שכאשר תגיע דגימה חדשה, המודל ידע להתאים עבורה תוצאה נכונה ביותר.

דוגמה לאוברפיטינג ברגרסיה:





Overfitting starting here

עצי החלטה נוטים ל-overfit. המשמעות של כך היא שהעצים שלנו נעשים יותר ויותר ספציפיים עבור ה-training data.

הדרך שלנו להתמודד עם נטייה זו נקראת Pruning.

קיצוץ Pruning

בגישה זו עלינו לקצץ ענפים מהעץ על מנת להפוך אותו לקטן יותר, ובכך נקבל test error טוב יותר.

- אפשר להחליט מראש על גובה עץ מקסימלי לפי ניסוי וטעייה. ננסה גבהים שונים ונחשב עבור כל גובה את ה-validation accuracy הטוב ביותר. (ממש כמו בגרדיאנט דיסנט עם האלפא learning rate)
- אופציה נוספת נקראית **post pruning**: נעלה במעלה העץ מהעלים לכיוון השורש, עבור כל צומת נחליט האם להשאיר אותו או לקצץ אותו מהעץ. נחליט זאת על בסיס השאלה האם הפיצול לאטריביוט הנ״ל תורם לנו או לא (Chi Square Test*).
- אפשר גם לקצץ בזמן בניית העץ = לא ניצור עוד צמתים בנים.

:*Chi Square Test

זוהי בדיקה סטטיסטית שמטרתה להגיד לנו האם הפיצול לפי אטריביוט כלשהו מניבה עבורנו התפלגות שהיא רנדומלית לחלוטין או האם יש לה כוח חיזוי כלשהו. לכן, נבדוק אם פיצול לפי האטריביוט הנבחר מניב עבורנו התפלגות שהיא ממש רנדומלית. אם יש לי בצומת 100 אינסטנסים, אם נחלק אותה רנדומלית, נצפה שהפיצול הרנדומלי יניב בערך את אותו היחס שהיה באבא. האם הילדים שומרים על היחס בין הקלאסים שיש באבא – אם כן, נגזום, לא הרווחנו וודאות! ככל שאנחנו יותר רחוקים ממה שציפינו = החלוקה היא יותר רנדומית. אם אני רחוק מרנדום, נבצע את החלוקה. בדיקה זו מבצעים תוך כדי הבנייה של העץ - בודקים את כוח החיזוי.

הבדיקה מבוצעת באופן הבא:

- The test itself (assume Y can only take values of 0 \ 1):
- $P(Y=0) \approx \frac{\# \, Y=0 \; instances}{\# Instances}$
- Call D_f = number of instances where $x_j = f$
- p_f = number of instances where $x_j = f \;\&\; Y = 0$
- n_f = number of instances where $x_j = f \;\&\; Y = 1$
- $E_0 = D_f * P(Y=0)$, $E_1 = D_f * P(Y=1)$
- So Chi Square statistic is:

$$\chi^2 = \sum_{f \in values(x_j)} \frac{\left(p_f - E_0\right)^2}{E_0} + \frac{\left(n_f - E_1\right)^2}{E_1}$$

עבור כל ילד נחשב ונסכום – כמה דגימות יש לנו בקלאס0 פחות כמות הדגימות שנצפה לקבל בקלאס0 בריבוע, לחלק לצפי הדגימות בקלאס0. ועוד מספר הדגימות בקלאס1 פחות כמות הדגימות שצפינו שנקבל בקלאס1 בריבוע, לחלק לצפי הדגימות בקלאס1. (החלוקה היא סתם נרמול)

המהות היא כמה אני קרוב לצפי שלי.

נקבל מספר כלשהו, ונשאל את עצמנו מה זה אומר? בשביל כך יש לנו את טבלת ה-Chi Square.

טבלת ה-Chi Square

מה שמעניין אותנו הם המספרים בתוך הטבלה.

נרצה להבין מה זה אומר Degree of rhythm.

DegreeOfRhythm = (numOfClasses – 1)(numOfValsInA – 1)

בכל צומת שאנו נמצאים יש כמות אחרת של ערכים לאטריביוט הנבחר. כלומר יתכן שלאטריביוט "תחזית" יש 4 אפשרויות (מעונן, מעונן חלקית, שמשי, גשום), אבל בצומת הנוכחי שלנו אין אפשרות ל"תחזית" שמשית. לכן numOfValinA = 3. מפני שזה מספר הערכים האפשריים של האטריביוט בצומת הנוכחי.

מהו האלפא ריסק Alpha Risk?

אני רוצה להיות רחוק מרנדום בהסתברות מסויימת (השגיאה שאני מוכנה לסבול למשל 0.05)

Table of Probabilities for the Chi-Squared Distribution

Alpha Risk

df	0.995	0.990	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.25	0.01	0.005	0.001
1	0.000039	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.323	2.706	3.841	1.323	6.635	7.879	10.828
2	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	2.773	9.210	10.597	13.816
3	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	4.108	11.345	12.838	16.266
4	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	5.385	13.277	14.860	18.467
5	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	6.626	15.086	16.750	20.515
6	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	7.841	16.812	18.548	22.458
7	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	9.037	18.475	20.278	24.322
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	10.219	20.090	21.955	26.124
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	11.389	21.666	23.589	27.877
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	12.549	23.209	25.188	29.588
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	13.701	24.725	26.757	31.264
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	14.845	26.217	28.300	32.909
13	3.565	4.107	5.009	5.892	7.042	9.299	12.340	15.984	19.812	22.362	15.984	27.688	29.819	34.528
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	17.117	29.141	31.319	36.123
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	18.245	30.578	32.801	37.697
16	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	19.369	32.000	34.267	39.252
17	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	20.489	33.409	35.718	40.790
18	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	21.605	34.805	37.156	42.312
19	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	22.718	36.191	38.582	43.820
20	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	23.828	37.566	39.997	45.315
21	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	24.935	38.932	41.401	46.797
22	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	26.039	40.289	42.796	48.268
23	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	27.141	41.638	44.181	49.728
24	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	28.241	42.980	45.559	51.179
25	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	29.339	44.314	46.928	52.620
26	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.435	35.563	38.885	30.435	45.642	48.290	54.052
27	11.808	12.879	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	31.528	46.963	49.645	55.476
28	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	32.620	48.278	50.993	56.892
29	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	33.711	49.588	52.336	58.301
30	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	34.800	50.892	53.672	59.703
40	20.707	22.164	24.433	26.509	29.051	33.660	39.335	45.616	51.805	55.758	45.616	63.691	66.766	73.402
50	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	56.334	76.154	79.490	86.661
60	35.534	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	66.981	88.379	91.952	99.607
70	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	77.577	100.425	104.215	112.317
80	51.172	53.540	57.153	60.391	64.278	71.145	79.334	88.130	96.578	101.879	88.130	112.329	116.321	124.839
90	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.565	113.145	98.650	124.116	128.299	137.208
100	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.141	118.498	124.342	109.141	135.807	140.169	149.449

דוגמה לחישוב האנטרופיה ולאחר מכן הכנסה לנוסחת ה-Information Gain:

- What calculations are needed to find the feature to split the root of the decision tree using Information Gain
- Reminder:

$$Information\_Gain = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Entropy(S) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} log \frac{|S_i|}{|S|}$$

  - c – number of classes
  - Values(A) – all the values in the A feature
- We need to calculate:
  - Entropy(root)
  - Weighted average of the Entropy according to “Attraction”
  - Weighted average of the Entropy according to “Weather”

Instance	Attraction	Weather	Classification
1	Swim	Hot	-
2	Dance	Hot	+
3	Casino	Hot	+
4	Golf	Hot	-
5	Swim	Mild	-
6	Casino	Mild	-
7	Dance	Mild	+
8	Golf	Mild	-
9	Ski	Mild	+
10	Ski	Cold	+
11	Casino	Cold	-
12	Dance	Cold	-

- Entropy(root)

$$Entropy(root) = -\left(\frac{7}{12} log \frac{7}{12} + \frac{5}{12} log \frac{5}{12}\right)$$

- Weighted average of the Entropy according to “Attraction”

$$\sum_{v \in Values(\text{Attraction})} \frac{|S_v|}{|S|} Entropy(S_v)$$
$$= -\left(\frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right) + \frac{3}{12}\left(\frac{2}{3} log \frac{2}{3} + \frac{1}{3} log \frac{1}{3}\right) + \frac{3}{12}\left(\frac{1}{3} log \frac{1}{3} + \frac{2}{3} log \frac{2}{3}\right) + \frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right) + \frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right)\right)$$

- Weighted average of the Entropy according to “Weather”

$$\sum_{v \in Values(\text{Weather})} \frac{|S_v|}{|S|} Entropy(S_v)$$
$$= -\left(\frac{4}{12}\left(\frac{2}{4} log \frac{2}{4} + \frac{2}{4} log \frac{2}{4}\right) + \frac{5}{12}\left(\frac{2}{5} log \frac{2}{5} + \frac{3}{5} log \frac{3}{5}\right) + \frac{3}{12}\left(\frac{1}{3} log \frac{1}{3} + \frac{2}{3} log \frac{2}{3}\right)\right)$$

Put it all together in the Information Gain formula

$$Information\ Gain(root, \text{Attraction})$$
$$= -\left(\frac{7}{12} log \frac{7}{12} + \frac{5}{12} log \frac{5}{12}\right)$$
$$+ \left(\frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right) + \frac{3}{12}\left(\frac{2}{3} log \frac{2}{3} + \frac{1}{3} log \frac{1}{3}\right) + \frac{3}{12}\left(\frac{1}{3} log \frac{1}{3} + \frac{2}{3} log \frac{2}{3}\right)\right.$$
$$\left. + \frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right) + \frac{2}{12}\left(\frac{2}{2} log \frac{2}{2}\right)\right) = 0.36$$

$$Information\ Gain(root, \text{Weather})$$
$$= -\left(\frac{7}{12} log \frac{7}{12} + \frac{5}{12} log \frac{5}{12}\right)$$
$$+ \left(\frac{4}{12}\left(\frac{2}{4} log \frac{2}{4} + \frac{2}{4} log \frac{2}{4}\right) + \frac{5}{12}\left(\frac{2}{5} log \frac{2}{5} + \frac{3}{5} log \frac{3}{5}\right)\right.$$
$$\left. + \frac{3}{12}\left(\frac{1}{3} log \frac{1}{3} + \frac{2}{3} log \frac{2}{3}\right)\right) = 0.0085$$