تصمیم‌گیری

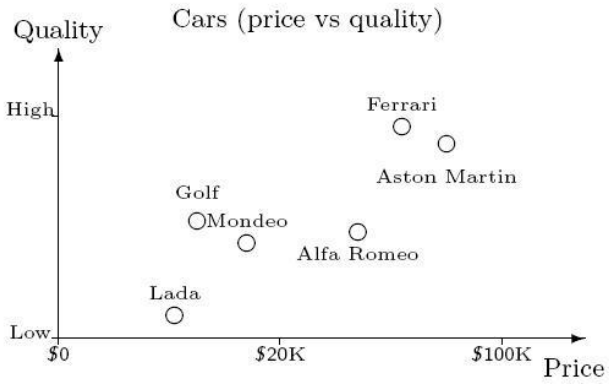
پرسش: تصمیم چیست؟ تصمیم‌گیری به چه معناست؟ تصمیم‌گیری چه نقشی در طراحی سیستم‌های خبره دارد؟

تابع سودمندی (Utility Function)

$$a \gtrsim b \quad \text{if and only if} \quad u(a) \geq u(b)$$

تصمیم‌گیری چند معیاره (Multi-criteria Decision Making)

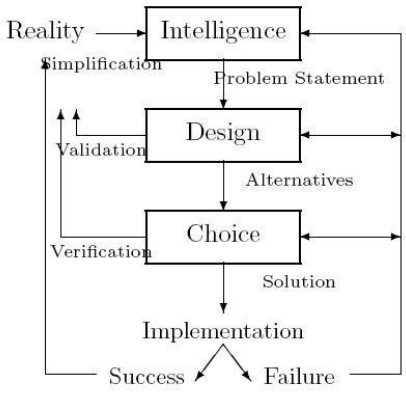
چندین معیار در تصمیم‌گیری، تاثیر دارد.



شکل10-1: مثال انتخاب خودرو

پرسش: در تصمیم گیری چندمعیاره، تابع سودمندی چگونه تعریف می شود؟

سه فاز مهم در تصمیم‌گیری (از نگاه سایمون)



شکل10-2: سه فاز تصمیم‌گیری

چون در انجام کاری مردد باشی، آن طرف را انتخاب کن که بی آزارتر است. (سعدی)

درخت‌تصمیم‌گیری (Decision Tree)

در بسیاری از مواقع اکتساب دانش از فرد خبره سخت و مشکل است؛ چون وقتی میزان خبرگی بالا می‌رود، فرد خبره، مسایل و جزئیاتی را بدیهی فرض‌می‌کند، درحالی‌که مهندس دانش نیاز به این جزئیات، برای طراحی سیستم خبره دارد. به عبارت ساده، در این موارد فرد خبره با سوال و جواب، پاسخ‌های مناسبی به مهندس دانش، نمی‌دهد. راه حل یا راه حل‌ها کدامند؟

در این موارد فرد خبره نمی‌تواند چگونگی حل مساله را متناسب با نیاز یک مهندس داش بیان کند. مهندس دانش خود باید راه حل مساله را با روش استقرا، بیابد.

استقراء (Induction)

ایجاد قوانین کلی از مجموعه مثال‌هاست. (مثال‌ها در قالب جداولی نمایش‌داده‌می‌شود که به آنها، جداول‌تصمیم‌گیری، گفته‌می‌شود.) از طریق مثال‌ها، دانش استخراج‌می‌شود و به همین دلیل استقرا یک نوع یادگیری است.

دانش استخراج شده درقالب درختی به نام، درخت تصمیم‌گیری، بازنمایی‌خواهدشود.

نکته: درخت تصمیم‌گیری؛ روشی برای نمایش یک سری از قوانین هستند که منتهی به یک رده یا مقدار می‌شوند.

روند کلی: از روی جدول تصمیم‌گیری(مثال‌ها)، درخت تصمیم‌گیری، ساخته‌شود و از روی درخت، قوانین‌کلی استخراج می‌شود. برای ایجاد درخت تصمیم‌گیری، از مفهوم آنتروپی[1] استفاده‌می‌شود.

پرسش: اگر سکه‌ای، یک میلیون بار بیندازیم، چنانچه این نتایج را بخواهیم برای فرد دیگری ارسال کنیم، به چه تعداد بیت نیاز خواهد بود؟

برای پاسخ به این سوال دو حالت زیر را در نظر بگیرید: 1) سکه سالم باشد 2) سکه ناسالم باشد

حالت اول : **سکه سالم باشد**

احتمال وقوع خط = احتمال وقوع شیر (در هر پرتاب)

تعداد بیتها جهت نمایش وضعیت هر پرتاب = 1 بیت

تعداد کل بیتهای مورد نیاز = 000 bits ‘ 1000

حالت دوم: **سکه ناسالم باشد**

فرض کنید احتمال شیرآمدن 1/1000 و احتمال خط آمدن 999/1000 باشد.

در یک میلیون پرتاب انتظار داریم 1000 بار شیر بیاید.

پرسش: چگونه راه حل را می‌توان بهتر کرد؟

نکته: دنباله نمایش سکه های ناسالم، حاوی اطلاعات کمتری نسبت به دنباله ی نمایش سکه های سالم است، لذا به بیتهای کمتری برای انتقال نیاز دارد.

محتوای اطلاعاتی (Information Content): تعداد بیت‌های لازم برای ارسال یک دنباله با استفاده از یک کدگذاری بهینه

همیشه می‌توان از یک کدگذاری که کمتر بهینه است، استفاده‌کرد. اما افزایش بیت‌ها هرگز به معنای افزایش محتوای اطلاعاتی نیست.

[1] - بیشتر مطالب مربوط به مبحث آنتروپی سری دهم (از صفحه ۶۴ تا ۷۰) برگرفته از اسلایدهای درس مباحث پیشرفته در پایگاه داده‌هاست که توسط دکتر میرمحسن پدارم ارئه‌شده است.

مثال: 1000 بار پرتاب یک تاس 8 وجهی را در نظر بگیرید:

الف) تاس سالم باشد:

تعداد بیتهای مورد نیاز برای بیان وضعیت پرتاب تاس : 3 بیت

به ارسال 3000 بیت نیاز خواهیم داشت.

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\frac{1}{64}+\frac{1}{128}+\frac{1}{128}=1$$

A B C D E F G H

ب) تاس ناسالم باشد: فرض کنید احتمال آمدن وجه $i(i=1 \dots 8)$ به صورت زیر باشد:

فرض کنید کدهای زیر را به هر یک از 8 حالت تاس اختصاص دهیم:

A ⟶ 0
B ⟶ 10
C ⟶ 110
D ⟶ 1110
E ⟶ 11110
F ⟶ 111110
G ⟶ 1111110
H ⟶ 1111111

نرخ وقوع آن حالت * تعداد بیتهای هر حالت $\sum$ = متوسط بیتهای مورد نیاز برای ارسال

آنتروپی

فرض کنید دو نفر X و Y وجود دارند که یکی از آنها یک سوال ساده می پرسد و دیگری به آن سوال پاسخ بله یا خیر می‌دهد.

مقدار اطلاعاتی که در این سوال-جواب‌ها بین طرفین رد و بدل می‌شود، آنتروپی نامیده‌می‌شود.

فکر می کنید با یک سوال-جواب چه مقدار اطلاعات بدست خواهید آورد؟

مثال: توپی داخل یکی از دو جعبه قرار دارد. الف : آیا توپ داخل جعبه سمت راست است؟ ب : بله

با یک سوال-جواب می توان فهمید که توپ کجاست. پس می توان گفت که **مقدار اطلاعات 1 بیت است.**

مثال: الف: آیا توپ داخل جعبه های بالایی است؟ ب : خیر

الف: آیا توپ داخل جعبه سمت راستی است؟ ب : خیر

مقدار اطلاعات 2 بیت است.

پرسش: صفر بیت چه چیزی را تعبیرمی کند؟ صفر بیت اطلاعات دریافت کرده‌اید؟؟؟!!!!!

کلاس‌بندی و آنتروپی

فرض کنید 5 کلاس A,B,C,D,E را داریم و می خواهیم یک شی را در آنها کلاس بندی کنیم.

الف) اگر عدم قطعیت نداشته باشیم:

	A	B	C	D	E
احتمال	0	1	0	0	0

$H(X) = 0\text{Log}(1/0) + 1\ \text{Log}(1/1) + 0\ \text{Log}(1/0) + \dots$

$= 0$ (عدم قطعیت نداریم)

ب) اگر بین دو کلاس شک داشته باشیم:

	A	B	C	D	E
احتمال	0.5	0.5	0	0	0

$H(X) = 0.5\text{Log}(1/0.5) + 0.5\text{Log}(1/0.5) + 0 \dots = 1$

یک بیت عدم قطعیت داریم، اگر اطلاعات کلاس را بخواهیم ارسال کنیم به یک بیت نیاز خواهیم داشت.

میان همه‌ی چیزهای قطعی، قطعی‌ترین شان تردید است. (برتولت برشت)

ج) اگر بین چهار کلاس شک داشته باشیم:

	A	B	C	D	E
احتمال	0.25	0	0.25	0.25	0.25

$H(X) = 0.25\text{Log}(1/0.25) + 0\text{Log}(1/0) + \ldots \quad = 2$

د) اگر حالت زیر را بین کلاس‌ها داشته باشیم:

	A	B	C	D	E
احتمال	0.5	0.125	0.125	0.125	0.125

$H(X) = 2.$

اگر بخواهیم اطلاعات را ارسال کنیم به طور متوسط (زمانی) به دو بیت احتیاج داریم.

A = 1 B = 000 , C = 001 , D = 010 , E = 011

نکته: هر چه همه‌ی کلاس‌ها به طوریکنواخت محتمل شوند،H(X) بزرگتر می‌شود؛ یعنی میزان عدم قطعیت در انتخاب کلاس، افزایش‌می‌یابد. **آنتروپی(عدم قطعیت)** و **اطلاعات** دو روی یک سکه اند! در تئوری اطلاعات تنها سمبلهایی که برای گیرنده دارای عدم قطعیت هستند، به عنوان اطلاعات قلمداد می‌شوند. (تنها اطلاعاتی که برای فهمیدن اساسی هستند باید فرستاده شوند.)

مفهوم اطلاعات

تئوری اطلاعات محتوای اطلاعات را بر حسب **بیت** اندازه گیری می‌کند. وقتی می‌خواهیم در مورد محتوای اطلاعاتی یک چیز صحبت کنیم مثل اینست که با چند پاسخ آری یا خیر وضعیت آن را مشخص‌کنیم. وضعیت یک سکه دو برآمد دارد پس با یک بیت می‌توانیم وضعیت آن را مشخص کنیم. اگر در مورد سکه 0.99 احتمال شیر باشد پس تقریبا همیشه وضعیت شیر رخ میدهد لذاسوال و جوابی نیاز نیست و محتوای اطلاعاتی به صفر نزدیک می‌شود.

بهره‌ی اطلاعاتی(Information Gain)

بعد از آزمون یک صفت به چند بیت اطلاعات دیگر نیازمندیم؟ صفت مورد آزمایش چقدر به ما اطلاعات می دهد؟

Gain (A) = Info (I) - Reminder (A) = Info (I) - Info (I,A)

نکته: در استفاده از بهره‌ی اطلاعاتی باید آن صفت یا Attribute ای را انتخاب‌کرد که بیشترین بهره‌ی اطلاعاتی را داراست و معمولا از فرمول نرمال‌سازی‌شده در محاسبه بهره اطلاعاتی استفاده می‌شود.

$$GainRatio(A) = \frac{Info(I) - Info(I,A)}{SplitInfo}$$

Split(A) عدم قطعیت نمونه ها نسبت به صفت A است.

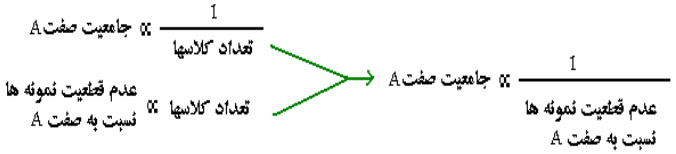
Split(A) در دو حالت بزرگ می شود:

1- تعداد شاخه های K(مقادیر صفت A) بیشتر شود.

2- تعداد نمونه ها در هر کلاس j کم باشد.

بروز دو حالت فوق موجب جامعیت کمتر صفت می‌شود.

$$Split(A) = -\sum_{j=1}^{k} \frac{\#classj}{\#I} \log_2 \frac{\#classj}{\#I}$$



نقطه آغاز تصمیم های بهتر، نقطه پایان تصمیم های بدتر است. (برایان تریسی)

پرسش: چرا از فرمول بهره‌ی اطلاعاتی به صورت نرمال شده استفاده می‌کنیم؟

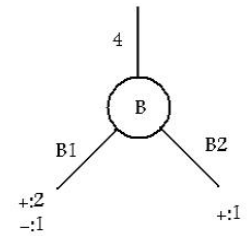
ما دنبال صفت Aای هستیم که: اولا بهره ی اطلاعاتی بالاتری داشته باشد. ثانیا جامعیت کلاس‌بندی آن بیشتر باشد.

مثال: فرض کنید صفت خروجی برای 5 ویژگی ورودی به صورت زیر باشد. مقدار Gain و Gain Ratio را برای هر صفت بدست آورید.

صفت A	صفت B	صفت C	صفت D	صفت E	صفت خروجی
					+
					+
					+
					−

ابتدا عدم قطعیت نمونه ها نسبت به صفت خروجی را بدست می‌آوریم

$$H(I) = -\frac{3}{4}\log_2^{\frac{3}{4}} - \frac{1}{4}\log_2^{\frac{1}{4}} = 0.81bit$$



$$H(I,B) = \frac{3}{4}H(B1) + \frac{1}{4}H(B2)$$

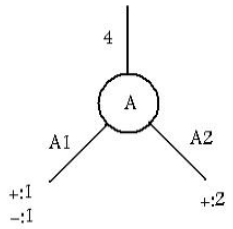
$$H(B1) = -\frac{2}{3}\log_2^{\frac{2}{3}} - \frac{1}{3}\log_2^{\frac{1}{3}} = 0.9183$$

$$H(B2) = 0bit$$

$$Gain(B) = H(I) - H(I,B) \approx 0.1212$$

$$Split(B) = -\frac{3}{4}\log_2^{\frac{3}{4}} - \frac{1}{4}\log_2^{\frac{1}{4}} = 0.81$$

$$GainRatio(B) = \frac{0.1212}{0.81} \approx 0.15$$



$$H(I,A) = \frac{2}{4}H(A1) + \frac{2}{4}H(A2)$$

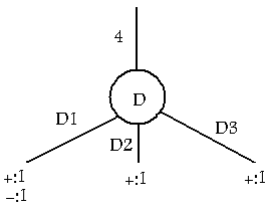
$$H(A1) = -\frac{1}{2}\log_2^{\frac{1}{2}} - \frac{1}{2}\log_2^{\frac{1}{2}} = 1bit$$

$$H(A2) = -\frac{2}{2}\log_2^{\frac{2}{2}} - \frac{0}{2}\log_2^{\frac{0}{2}} = 0bit$$

$$Gain(A) = H(I) - H(I,A) = 0.81 - \frac{1}{2}\times 1 = 0.31$$

$$Split(A) = -\frac{2}{4}\log_2^{\frac{2}{4}} - \frac{2}{4}\log_2^{\frac{2}{4}} = 1bit$$

$$GainRatio(A) = \frac{0.31}{1} = 0.31$$



$$H(I,D) = \frac{2}{4}H(D1) + \frac{1}{4}H(D2) + \frac{1}{4}H(D3)$$

$$H(D1) = 1bit$$

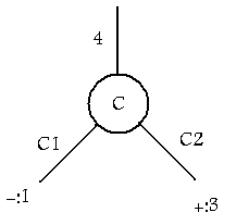
$$H(D2) = 0$$

$$H(D3) = 0$$

$$Gain(D) = 0.81 - 0.5 = 0.31$$

$$Split(D) = -\frac{2}{4}\log_2^{\frac{2}{4}} + \frac{1}{4}\log_2^{\frac{1}{4}} + \frac{1}{4}\log_2^{\frac{1}{4}} = 1.5$$

$$GainRatio(D) = \frac{0.31}{1.5} = 0.201$$



$$H(I,C) = \frac{3}{4}H(C1) + \frac{1}{4}H(C2)$$

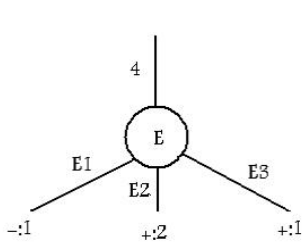
$$H(C1) = 0bit$$

$$H(C2) = 0bit$$

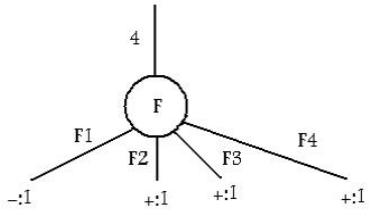
$$Gain(C) = H(I) - H(I,C) = 0.81$$

$$Split(C) = -\frac{1}{4}\log_2^{\frac{1}{4}} - \frac{3}{4}\log_2^{\frac{3}{4}} = 0.81$$

$$GainRatio(B) = \frac{1}{0.81} = 1.23$$



$$H(I,E)=\frac{1}{4}H(E1)+\frac{2}{4}H(E2)+\frac{1}{4}H(E3)$$
$$H(E1)=0$$
$$H(E2)=0$$
$$H(E3)=0$$
$$Gain(E)=0.81$$
$$Split(E)=-\frac{1}{4}\log_2^{\frac{1}{4}}-\frac{2}{4}\log_2^{\frac{2}{4}}-\frac{1}{4}\log_2^{\frac{1}{4}}=1.5$$
$$GainRatio(E)=\frac{0.81}{1.5}=0.54$$



$$H(I,F)=\frac{1}{4}H(F1)+\frac{1}{4}H(F2)+\frac{1}{4}H(F3)+\frac{1}{4}H(F4)$$
$$H(F1)=0$$
$$H(F2)=0$$
$$H(F3)=0$$
$$H(F4)=0$$
$$Gain(F)=0.81-0=0.81$$
$$Split(F)=-\frac{1}{4}\log_2^{\frac{1}{4}}-\frac{1}{4}\log_2^{\frac{1}{4}}-\frac{1}{4}\log_2^{\frac{1}{4}}-\frac{1}{4}\log_2^{\frac{1}{4}}=2$$
$$GainRatio(F)=\frac{0.81}{2}=0.405$$

اگر نقطه ی Gain مدنظر قرار گیرد: **C = F = E > A = D > B**

اگر GainRatio مدنظر قرار گیرد: **C > E > F > A > D > B**

مثال: استفاده از الگوریتم ID3 برای پیش بینی بارش باران بر اساس متغیرهای ورودی (فاکتورهای) زیر:

	Input attributes or Decision factors			result
	Sky	barometer	Wind	Rain
1	clear	rising	north	-
2	cloudy	rising	south	+
3	cloudy	steady	north	+
4	clear	falling	north	-
5	cloudy	falling	north	+
6	cloudy	rising	north	+
7	cloudy	falling	south	-
8	clear	rising	south	-

$$H(I)=-\sum_{i=1}^{2}\frac{\#classi}{\#I}\log_2^{\frac{\#classi}{\#I}}$$
$$=-\frac{4}{8}\log_2^{\frac{4}{8}}-\frac{4}{8}\log_2^{\frac{4}{8}}$$
$$=1bit$$

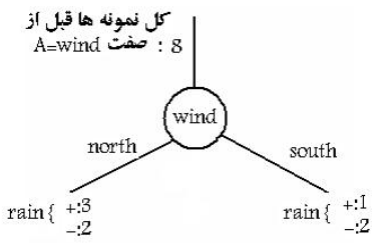
عدم قطعیت در کل نمونه ها

یا

تعداد بیتها برای بیان عدم قطعیت

حال با صفات (فاکتورهای) Sky, Barometer , wind مواجه‌ایم. برای هر کدام از این صفات باید Gain را محاسبه کنیم.

حالت اول: A=wind

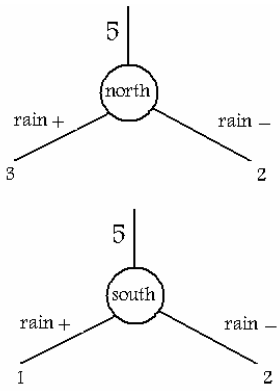


$$H(I,wind)=\sum_{i=1}^{2}\frac{\#classi}{\#I}H(classi)$$
$$=\frac{\#class(north)}{\#I}H(north)+\frac{\#class(south)}{\#I}H(south)$$
$$=\frac{5}{8}H(north)+\frac{3}{8}H(south)\quad ①$$

$$H(I,wind)=\frac{5}{8}H(north)+\frac{3}{8}H(south)$$
$$=\frac{5}{8}0.971+\frac{3}{8}0.918=0.951bits$$
$$Gain(wind)=H(I)-H(I\mid wind)$$
$$=1-0.951=0.049bits$$
$$Split(wind)=-\sum_{j=1}^{2}\frac{\#classj}{\#I}\log_2^{\frac{\#classj}{\#I}}$$
$$=-\frac{5}{8}\log_2^{\frac{5}{8}}-\frac{3}{8}\log_2^{\frac{3}{8}}=0.9544$$
$$GainRatio(wind)=\frac{0.049}{0.9544}=0.0513$$

$$H(north)=-\sum_{j=1}^{2}\frac{\#classj}{north}\log_2^{\frac{\#classj}{north}}$$
$$=-\frac{\#class+}{north}\log_2^{\frac{\#class+}{north}}-\frac{\#class-}{north}\log_2^{\frac{\#class+}{north}}$$
$$=-\frac{3}{5}\log_2^{\frac{3}{5}}-\frac{2}{5}\log_2^{\frac{2}{5}}=0.971bits \quad (2)$$

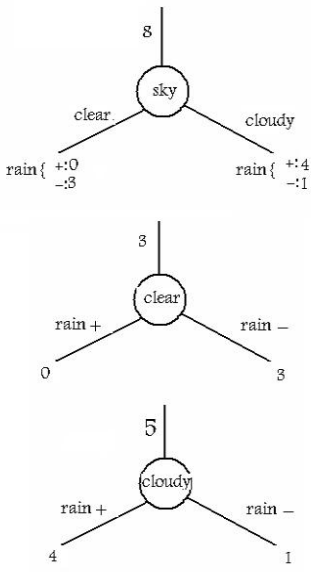
$$H(south)=-\frac{1}{3}\log_2^{\frac{1}{3}}-\frac{2}{3}\log_2^{\frac{2}{3}}=0.918bits \quad (3)$$



حالت دوم: A=sky

$$H(I,sky)=-\sum_{j=1}^{2}\frac{\#classj}{\#I}H(classj)$$
$$(1)' =\frac{3}{8}H(clear)+\frac{5}{8}H(cloudy)$$
$$H(clear)=-\frac{0}{3}\log_2^{\frac{0}{3}}-\frac{3}{3}\log_2^{\frac{3}{3}}=0bits \quad (2)'$$
$$H(cloudy)=-\frac{4}{5}\log_2^{\frac{4}{5}}-\frac{1}{5}\log_2^{\frac{1}{5}}=0.722bits \quad (3)'$$

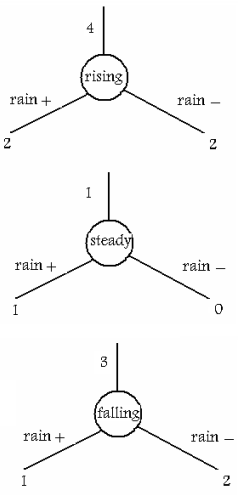
$$(1)',(2)',(3)' \quad H(I,sky)=-\frac{3}{8}\times 0+\frac{5}{8}\times 0.722$$
$$=0.45bits$$
$$Gain(sky)=H(I)-H(I\mid sky)$$
$$=1-0.45=0.55bits$$
$$Split(sky)=-\frac{3}{8}\log_2^{\frac{3}{8}}-\frac{5}{8}\log_2^{\frac{5}{8}}=0.9544$$
$$GainRatio(sky)=\frac{0.55}{0.9544}=0.5763$$

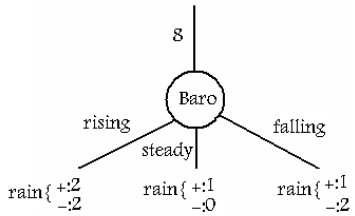


حالت سوم: A=Barometer

$$H(I,Baro)=-\sum_{j=1}^{3}\frac{\#classj}{\#I}H(classj)$$
$$(1)'' =\frac{4}{8}H(rising)+\frac{1}{8}H(steady)+\frac{3}{8}H(falling)$$

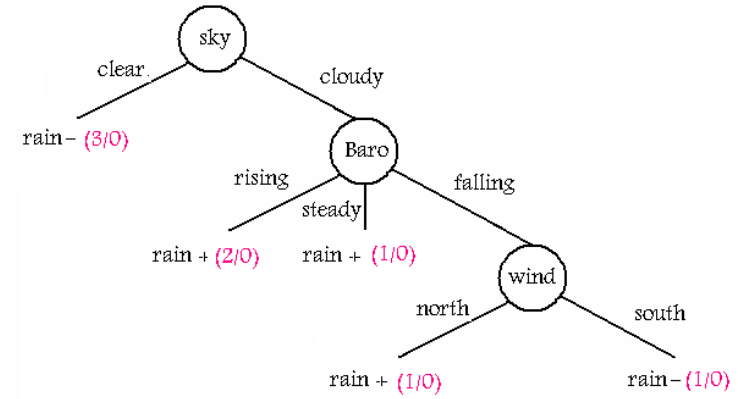
$$H(rising)=-\frac{2}{4}\log_2^{\frac{2}{4}}-\frac{2}{4}\log_2^{\frac{2}{4}}=1bits \quad (2)''$$
$$H(steady)=-\frac{1}{1}\log_2^{\frac{1}{1}}-\frac{0}{1}\log_2^{\frac{0}{1}}=0bits \quad (3)''$$
$$H(falling)=-\frac{1}{3}\log_2^{\frac{1}{3}}-\frac{2}{3}\log_2^{\frac{2}{3}}=0.9184bits \quad (4)''$$





$$H(I,Baro) = -\frac{4}{8}\times 1 + 0 + \frac{3}{8}\times 0.9182$$
$$= 0.8443 bits$$
$$Gain(Baro) = H(I) - H(I \mid Baro)$$
$$= 1 - 0.8443 = 0.1557 bits$$
$$Split(Barometer) = -\frac{4}{8}\log_2^{\frac{4}{8}} - \frac{1}{8}\log_2^{\frac{1}{8}} - \frac{3}{8}\log_2^{\frac{3}{8}}$$
$$= 1.406$$
$$GainRatio(Barometer) = \frac{0.1557}{1.406} = 0.1107$$

صفتی را که بیشترین بهره اطلاعاتی را دارد انتخاب‌می‌کنیم. در نهایت درخت تصمیم گیری به صورت زیر خواهد شد:



تمرین‌ها

تمرین10-1: اگر جدولی به صورت زیر داشته‌باشیم، در خت تصمیم گیری آن به چه صورت خواهدبود؟

A	B	C	خروجی
A1	B1	C1	+
A1	B2	C1	+
A2	B1	C2	-
A2	B2	C2	-

تمرین10-2: تصمیم گیری در سیستم‌های پشتیبان تصمیم گیر یا DSS چه نقشی دارد؟ نقش DSS در سیستم‌های خبره چیست؟

تمرین10-3*: الگوریتم C4.5 چه تفاوتی با الگوریتم ID3 دارد؟ الگوریتم C5.0 چه تفاوتی با الگوریتم C4.5 دارد؟

تمرین10-4*: الگوریتم Fuzzy ID3 چگونه عمل‌می‌کند و برای چه نوع مسائلی مناسب است؟