

درخت‌های تصمیم

سید ناصر رضوی n.razavi@tabrizu.ac.ir

۱۳۹۵

درخت‌های تصمیه و شبکه‌های عصبی

۲



□ بیان رسمی یادگیری.

□ سازگاری

□ سادگی

□ درخت‌های تصمیم.

□ قدرت بیان

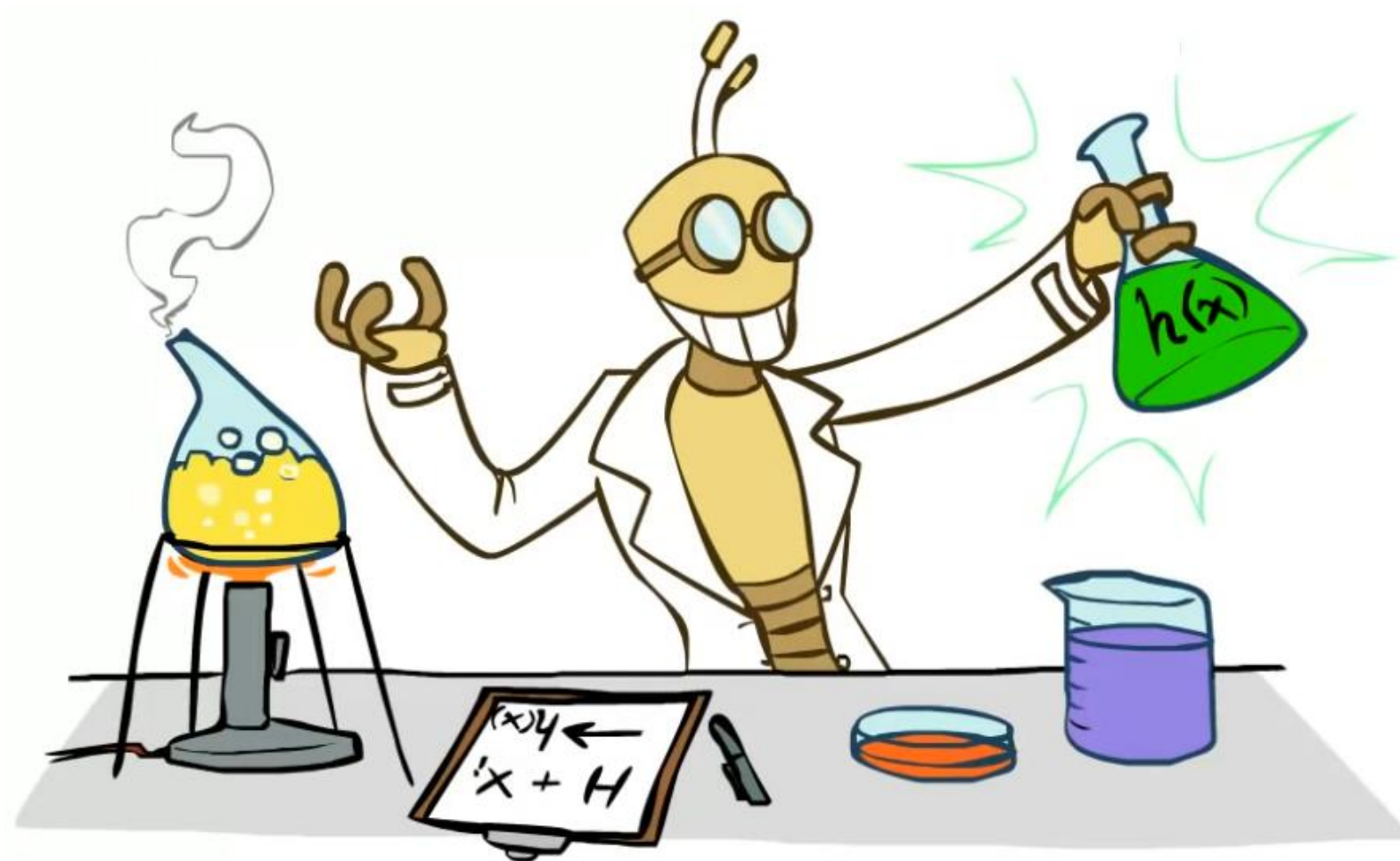
□ بهره اطلاعاتی

□ بیش برآزش

□ شبکه‌های عصبی.

یادگیری استقرایی

۴



یادگیری استقرایی

۵

□ ساده‌ترین شکل. یادگیری یک تابع از روی تعدادی نمونه

□ تابع هدف: g

□ نمونه‌ها: زوج مرتب‌های ورودی-خروجی $(x, g(x))$

□ مثلاً x ممکن است یک ایمیل باشد و $g(x)$ بیانگر معتبر بودن یا هرزنامه بودن آن

□ مثلاً x ممکن است یک خانه باشد و $g(x)$ بیانگر قیمت فروش آن خانه

□ مسئله.

□ با داشتن یک فضای فرضیه مانند H

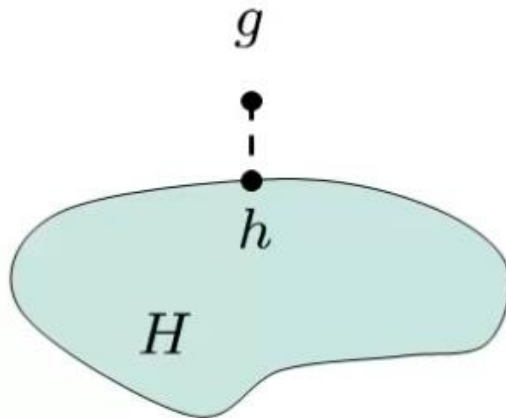
□ و یک مجموعه از نمونه‌های آموزشی مانند x_i

□ یک فرضیه مانند $h(x)$ پیدا کن به گونه‌ای که $h(x) \sim g(x)$

□ انواع

□ کلاس‌بندی (خروجی‌های چند مقداری)

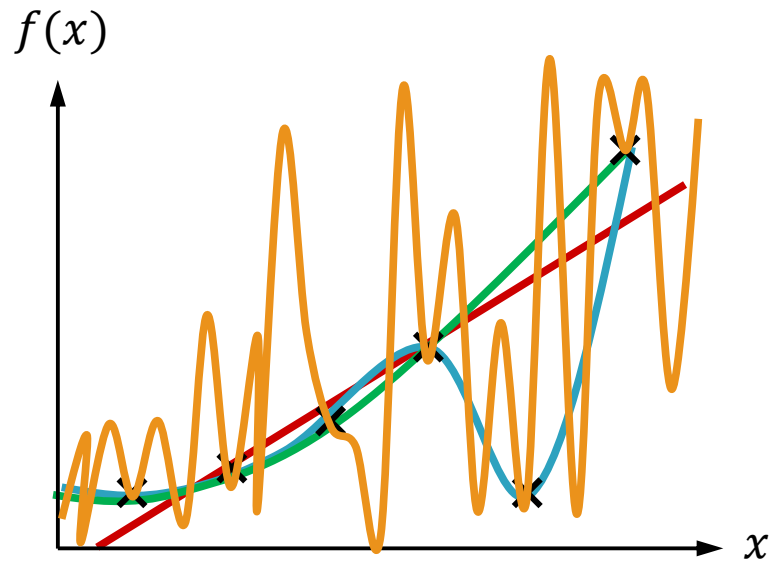
□ رگرسیون (خروجی‌های پیوسته)



یادگیری استقرایی

۶

□ برازش منحنی. [رگرسیون، تقریب تابع]



□ سازگاری در برابر سادگی.

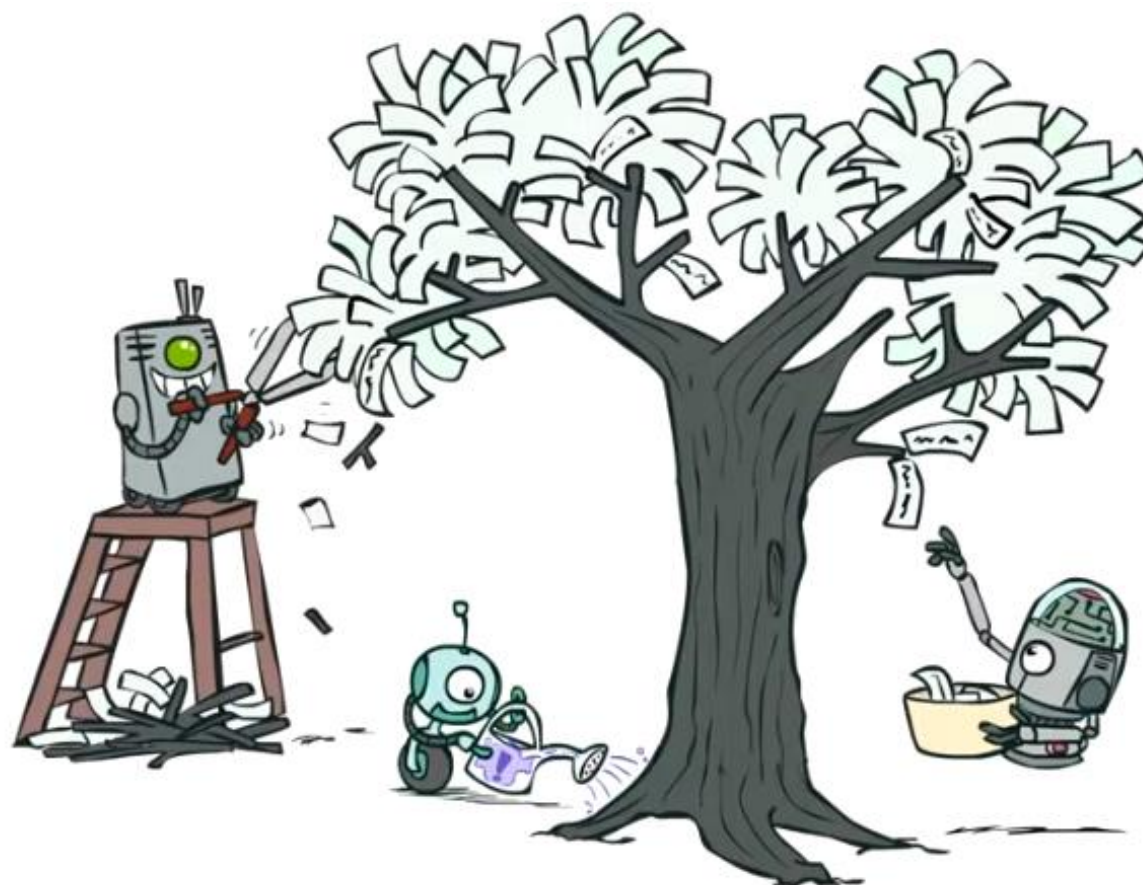
□ اصل تراش اوکام.

سازگاری در برابر سادگی

۷

- یک موازنه پایه‌ای. بایاس و واریانس
- معمولاً الگوریتم‌ها به صورت پیش فرض سازگاری را ترجیح می‌دهند [چرا؟]
- روش‌های مختلف به منظور عملیاتی کردن «سادگی».
- کاهش فضای فرضیه.
 - افزایش تعداد فرضیه‌ها: مانند فرض استقلال در بیز ساده
 - انتخاب ویژگی: داشتن ویژگی‌های کمتر با کیفیت بالاتر
 - محدودیت‌های ساختاری دیگر: لیست تصمیم در برابر درخت تصمیم
- تنظیم.
 - هموارسازی
 - بسیاری دیگر از پارامترهای تعمیم

درخت تصمیه



یادآوری: ویژگی‌ها

۹

Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X_1	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X_2	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X_3	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X_4	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X_5	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X_6	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X_7	F	T	F	F	None	\$	T	F	Burger	0-10	F
X_8	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X_9	F	T	T	F	Full	\$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X_{11}	F	F	F	F	None	\$	F	F	Thai	0-10	F
X_{12}	T	T	T	T	Full	\$	F	F	Burger	30-60	T

درخت‌های تصمیم

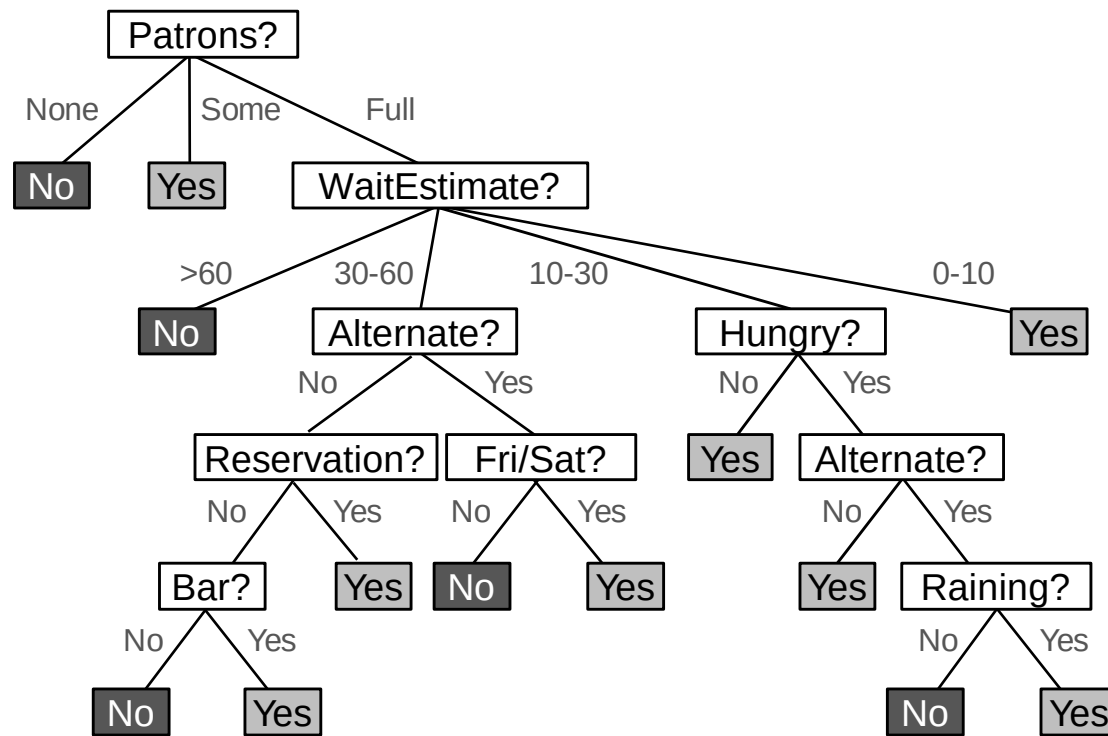
۱۰

□ نمایش تابع به شکل فشرده.

□ جدول درستی

□ جدول احتمالات شرطی

□ مقادیر رگرسیون



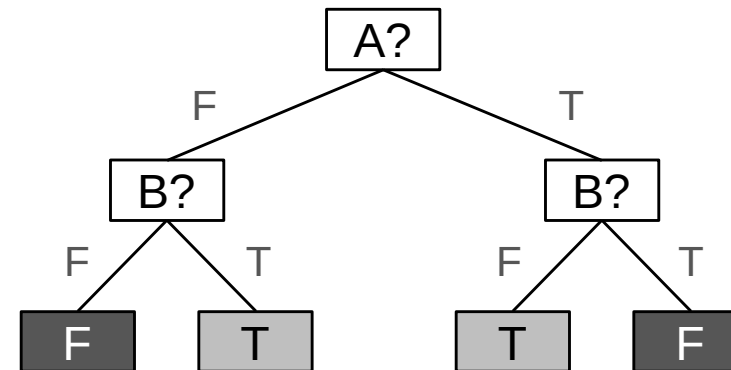
قدرت بیان درخت‌های تصمیم

۱۱

□ قابلیت بازنمایی هر تابع ممکن از ویژگی‌ها.

A	B	$A \text{ xor } B$
F	F	F
F	T	T
T	F	T
T	T	F

$$P(C|A, B)$$



□ هر چند ما درخت‌های فشرده‌تر را ترجیح می‌دهیم.

فضای فرضیه

۱۲

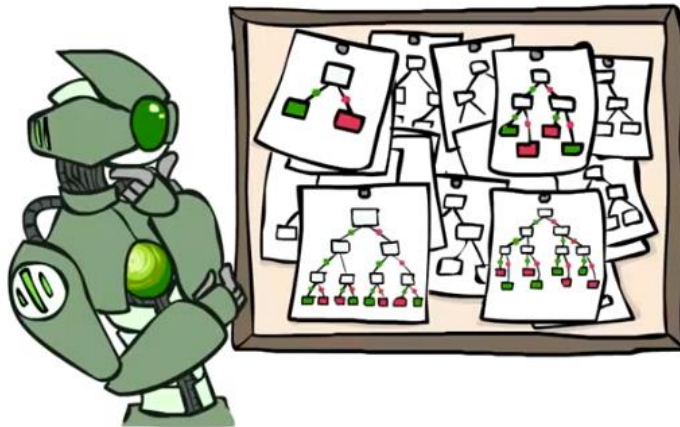
□ فضای فرضیه بزرگ.

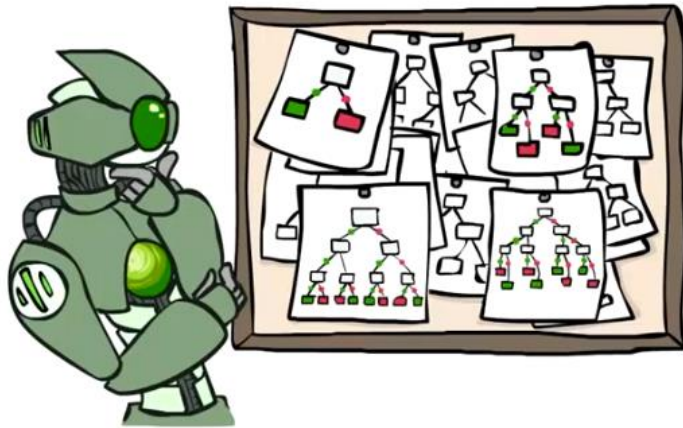
□ افزایش شانس برای نمایش تابع هدف.

□ افزایش تعداد فرضیه‌های سازگار با مجموعه آموزشی.

□ به این معنا که بالقوه می‌توانیم پیش‌بینی‌های بهتری انجام دهیم [بایاس کمتر]

□ اما در عمل ممکن است پیش‌بینی‌های بدتری به دست آوریم [واریانس بیشتر]





□ تعداد درخت‌های تصمیم متمایز با n ویژگی بولی.

□ = تعداد توابع بولی ممکن بر روی n ویژگی.

□ = تعداد جدول‌های درستی ممکن با 2^n سطر.

$$\square = 2^{(2^n)}$$

■ برای ۶ ویژگی، ۱۸۴۴۶۷۴۴۰۷۳۷۰۹۵۵۱۶۱۶ درخت!

□ تعداد درخت‌های تصمیم با عمق ۱.

□ = تعداد توابع بولی ممکن بر روی ۱ ویژگی.

□ = تعداد جدول‌های درستی ممکن با ۲ سطر ضرب در n .

$$\square = 4n$$

■ برای ۶ ویژگی، ۲۴ درخت!

یادگیری درخت تصمیم

۱۴

□ هدف. یافتن یک درخت کوچک سازگار با نمونه‌های آموزشی

□ ایده. انتخاب «مهم‌ترین ویژگی» برای قرار دادن به عنوان ریشه‌ی زیردرخت‌ها.

```
function DECISION-TREE-LEARNING(examples, attributes, parent_examples) returns  
a tree
```

```
if examples is empty then return PLURALITY-VALUE(parent_examples)
```

```
else if all examples have the same classification then return the classification
```

```
else if attributes is empty then return PLURALITY-VALUE(examples)
```

```
else
```

```
   $A \leftarrow \arg \max_{a \in \text{attributes}} \text{IMPORTANCE}(a, \text{examples})$ 
```

```
  tree  $\leftarrow$  a new decision tree with root test A
```

```
  for each value  $v_k$  of A do
```

```
    exs  $\leftarrow \{e : e \in \text{examples} \text{ and } e.A = v_k\}$ 
```

```
    subtree  $\leftarrow$  DECISION-TREE-LEARNING(exs, attributes – A, examples)
```

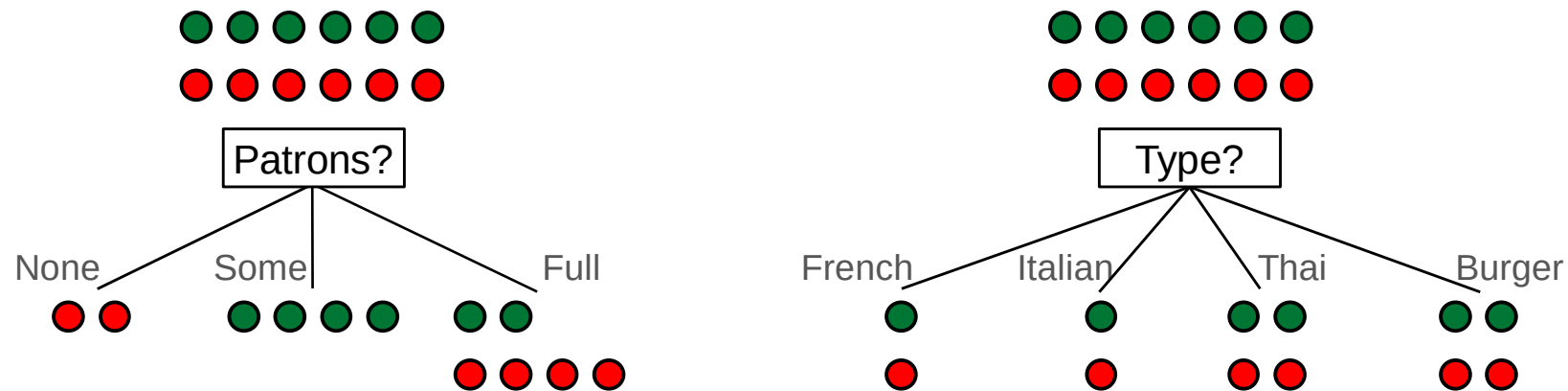
```
    add a branch to tree with label ( $A = v_k$ ) and subtree subtree
```

```
  return tree
```

انتخاب یک ویژگی

۱۵

□ ایده. یک ویژگی خوب به صورت ایده‌آل نمونه‌ها را به زیرمجموعه‌هایی تقسیم می‌کند که تمام نمونه‌های متعلق به آنها همگی یا «مثبت» هستند و یا «منفی».



□ در نتیجه به معیاری نیاز داریم که به وسیله‌ی آن بتوانیم «میزان خوب بودن» یک تقسیم‌بندی را اندازه‌گیری کنیم.

آنتروپی و اطلاعات

۱۶

□ اطلاعات به پرسش‌ها پاسخ می‌دهد.

□ هرچه در ابتدا میزان عدم قطعیت در رابطه با پاسخ بیشتر باشد، آنگاه اطلاعات بیشتری در پاسخ نهفته است. یعنی، اطلاعات از بین برنده‌ی عدم قطعیت است.

□ واحد اندازه‌گیری اطلاعات: بیت

■ پاسخ یک پرسش دو گزینه‌ای با احتمال پیشین $\langle 1/2, 1/2 \rangle$ ؟

■ پاسخ یک پرسش چهار گزینه‌ای با احتمال پیشین $\langle 1/4, 1/4, 1/4, 1/4 \rangle$ ؟

■ پاسخ یک پرسش چهار گزینه‌ای با احتمال پیشین $\langle 0, 0, 0, 1 \rangle$ ؟

■ پاسخ یک پرسش سه گزینه‌ای با احتمال پیشین $\langle 1/2, 1/4, 1/4 \rangle$ ؟

□ اگر احتمال برابر با p باشد، تعداد بیت‌های لازم برابر است با $\log \frac{1}{p}$.

- پاسخ کلی. اگر احتمال پیشین برابر با $\langle p_1, p_2, \dots, p_n \rangle$ باشد:
- اطلاعات برابر است با طول کد مورد انتظار.

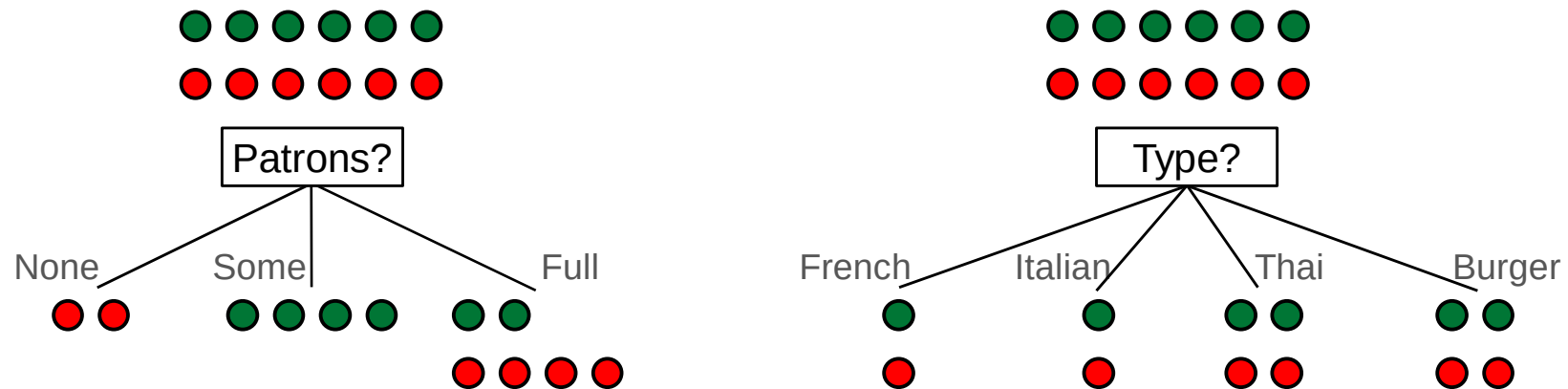
$$H(\langle p_1, p_2, \dots, p_n \rangle) = \sum_{i=1}^n -p_i \log_2 p_i$$

- کمیت بالا آنتروپی توزیع نام دارد.
- هرچه یک توزیع یکنواخت تر باشد، آنتروپی بالاتر است.
- هرچه یک توزیع دارای مقادیر بیشتری باشد، آنتروپی بالاتر است.
- هرچه یک توزیع تیزتر باشد، آنتروپی پایین تر است.

بهره اطلاعاتی

۱۸

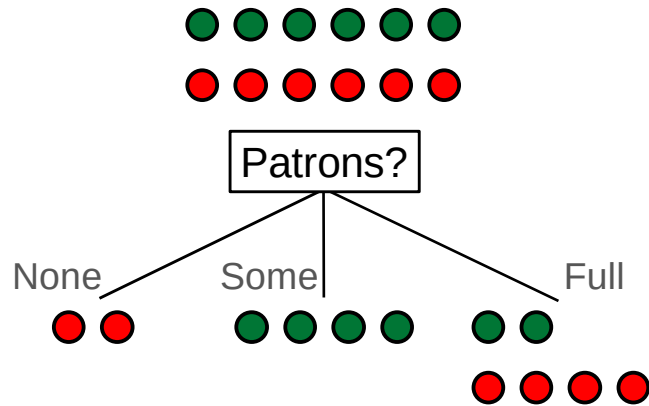
- بازگشت به درخت تصمیم!
- برای هر تقسیم، آنترپی را قبل و بعد از تقسیم مقایسه کن.
- این اختلاف **بهره‌ی اطلاعاتی** نام دارد.
- مشکل: پس از عمل تقسیم، بیش از یک توزیع وجود دارد!



- راه حل: استفاده از **آنترپی مورد انتظار**، وزن دهی شده با تعداد نمونه‌ها!

گام بعدی: فراخوانی بازگشتی

۱۹



- اکنون باید ساختن درخت را ادامه دهیم.
- کار دو شاخه به پایان رسیده است (چرا؟)
- برای شاخه‌ی **full** باید چه کار کنیم؟
- بین چه نمونه‌هایی آنجا وجود دارد ...

Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X_1	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X_2	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X_3	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X_4	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X_5	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X_6	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X_7	F	T	F	F	None	\$	T	F	Burger	0-10	F
X_8	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X_9	F	T	T	F	Full	\$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X_{11}	F	F	F	F	None	\$	F	F	Thai	0-10	F
X_{12}	T	T	T	T	Full	\$	F	F	Burger	30-60	T

گام بعدی: فراخوانی بازگشتی

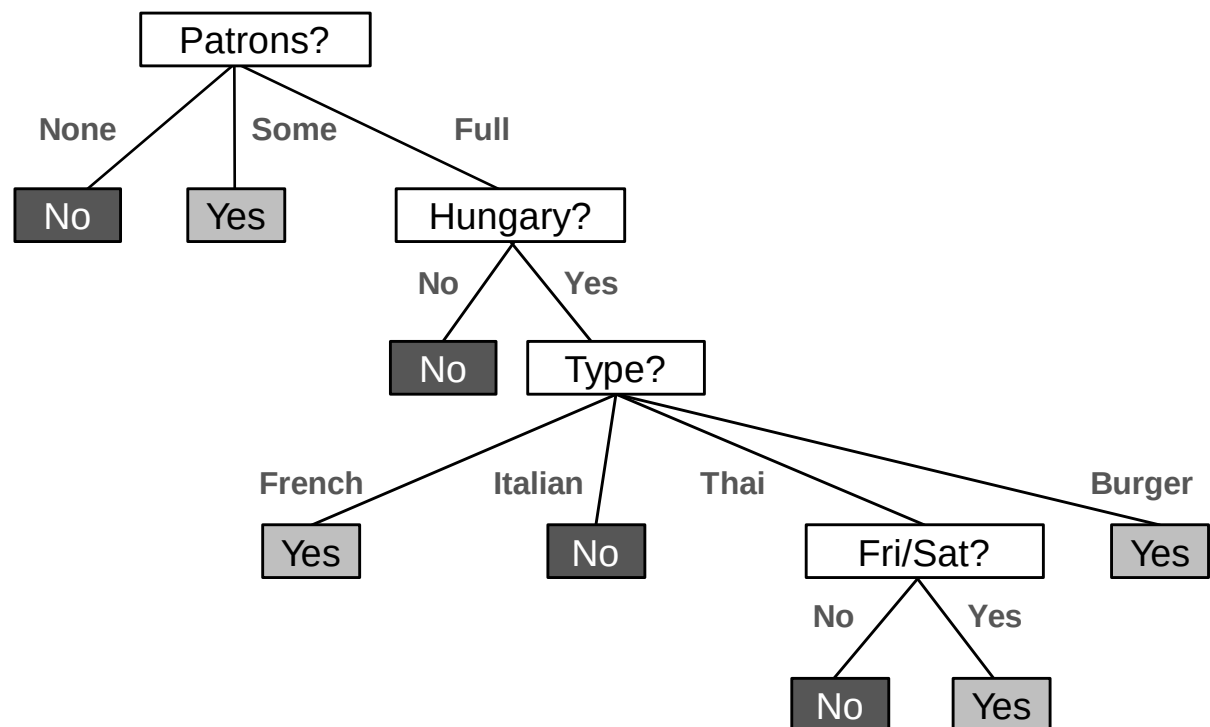
۲۰

Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X_1	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X_2	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X_3	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X_4	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X_5	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X_6	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X_7	F	T	F	F	None	\$	T	F	Burger	0-10	F
X_8	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X_9	F	T	T	F	Full	\$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X_{11}	F	F	F	F	None	\$	F	F	Thai	0-10	F
X_{12}	T	T	T	T	Full	\$	F	F	Burger	30-60	T

مثال: درخت تصمیم

۲۱

□ مثال. درخت تصمیم یاد گرفته شده از این ۱۲ نمونه.



□ از درخت «واقعی» بسیار ساده‌تر است.

□ این درخت اگرچه معقول است، اما صحیح نیست!