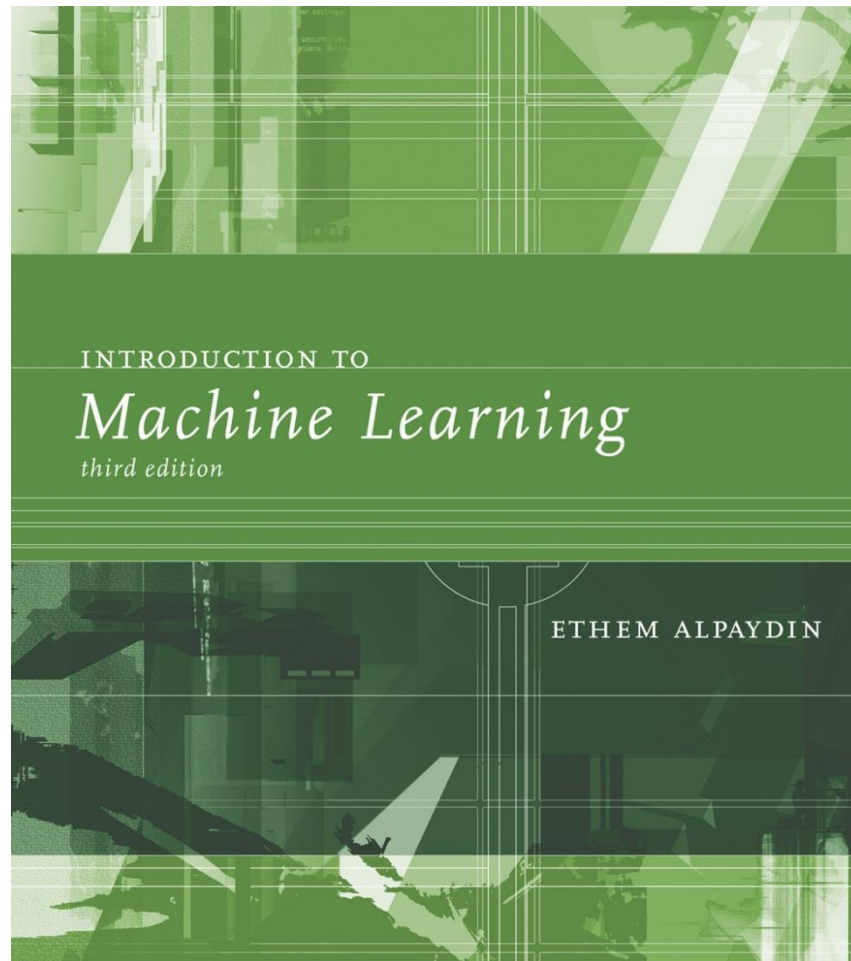


یادگیری ماشین: معرفی

سید ناصر رضوی n.razavi@tabrizu.ac.ir

۱۳۹۷

منابع و مراجع (۱)



□ مقدمه‌ای بر یادگیری ماشین.

[آلپایدین، ویراست سوم؛ ۲۰۱۴]

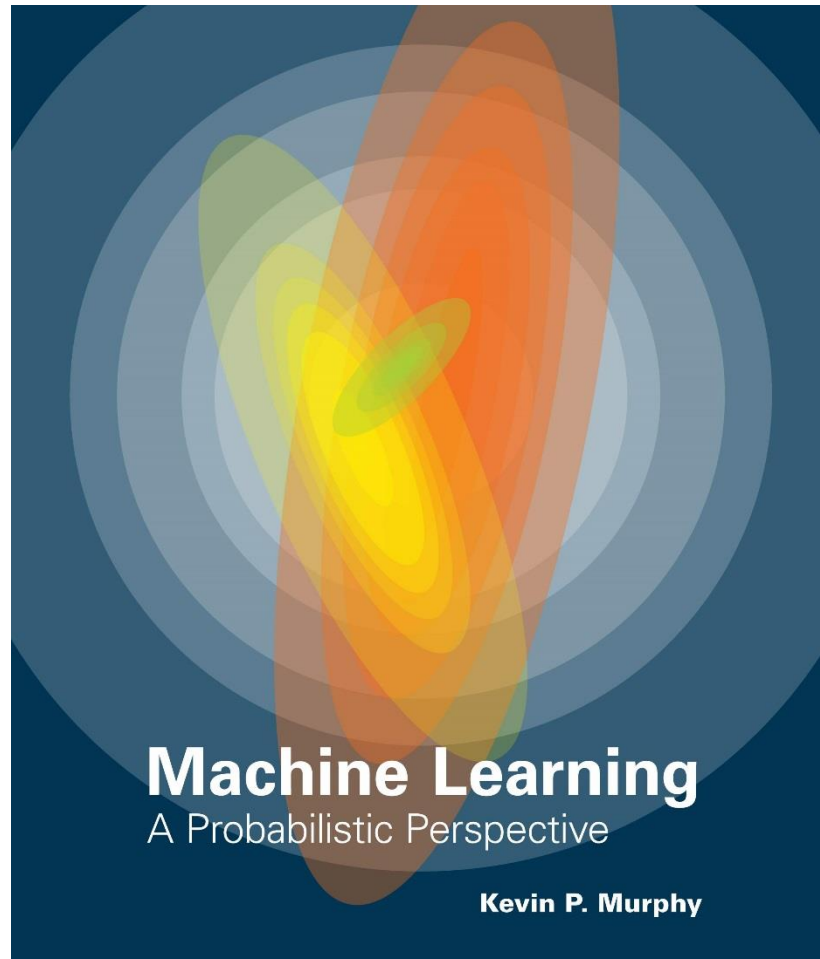
□ یادگیری ماشین: یک دیدگاه احتمالاتی.

[کونین مورفی، ۲۰۱۲]

□ شناسایی الگو و یادگیری ماشین.

[کریستوفر بیشاپ، ۲۰۰۶]

منابع و مراجع (۲)



□ مقدمه‌ای بر یادگیری ماشین.

[آلپایدین، ویراست سوم؛ ۲۰۱۴]

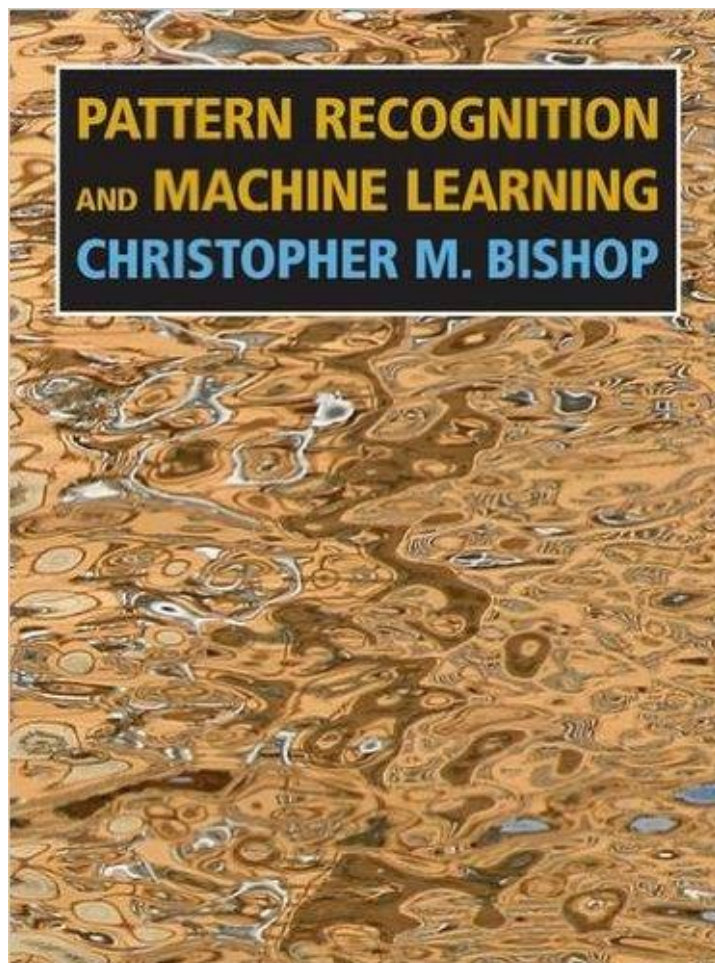
□ یادگیری ماشین: یک دیدگاه احتمالاتی.

[کون مورفی، ۲۰۱۲]

□ شناسایی الگو و یادگیری ماشین.

[کریستوفر بیشاپ، ۲۰۰۶]

منابع و مراجع (۳)



□ مقدمه‌ای بر یادگیری ماشین.

[آلپایدین، ویراست سوم؛ ۲۰۱۴]

□ یادگیری ماشین: یک دیدگاه احتمالاتی.

[کونین مورفی، ۲۰۱۲]

□ شناسایی الگو و یادگیری ماشین.

[بیشاپ، ۲۰۰۶]

- روش‌های تحلیل و طراحی الگوریتم‌ها
- تحلیل پیچیدگی محاسباتی الگوریتم‌های یادگیری

□ جبر خطی

- ماتریس‌ها، بردارها، عملیات ماتریسی و دستگاه معادلات خطی
- ماتریس وارون، بردارهای ویژه، مرتبه ماتریس، تجزیه مقادیر منفرد

□ حساب چند متغیره

- مشتق، انتگرال، صفحات مماس

□ احتمالات

- متغیرهای تصادفی، مقدار مورد انتظار، واریانس و ...



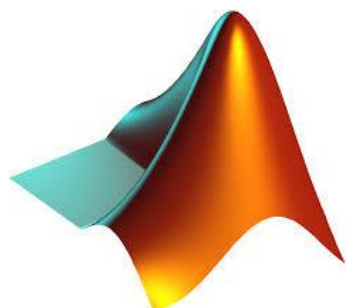
□ تمرین‌ها [۴۰٪]

□ مباحث نظری

□ برنامه‌نویسی

□ امتحان پایان‌ترم [۵۰٪]

□ حضور مؤثر در کلاس [۱۰٪]



MATLAB



OCTAVE

فهرست مطالب

- یادگیری نظارت شده.
- رگرسیون – رگرسیون خطی تک متغیره و چند متغیره
- دسته‌بندی – رگرسیون لجستیک، شبکه‌های عصبی، ماشین‌های بردار پشتیبان
- یادگیری بدون نظارت.
- خوشه‌بندی
- یادگیری تقویتی.
- برنامه‌نویسی با استفاده از زبان پایتون (یا اکتاو).
- توصیه‌های عملی در استفاده از الگوریتم‌های یادگیری ماشین.

چند نقل قول

«هر گام رو به جلو در جهت یادگیری ماشین ده‌ها برابر مایکروسافت ارزش دارد»

بیل گیتس - مدیر مایکروسافت



«نسل بعدی اینترنت چیزی به جز یادگیری ماشین نیست»

تونی تدر - مدیر اسبق دارپا



«یادگیری ماشین در نهایت به یک انقلاب واقعی منجر خواهد شد»

گرگ پاپادوپولس - مدیر اسبق سان



یادگیری ماشین چیست؟

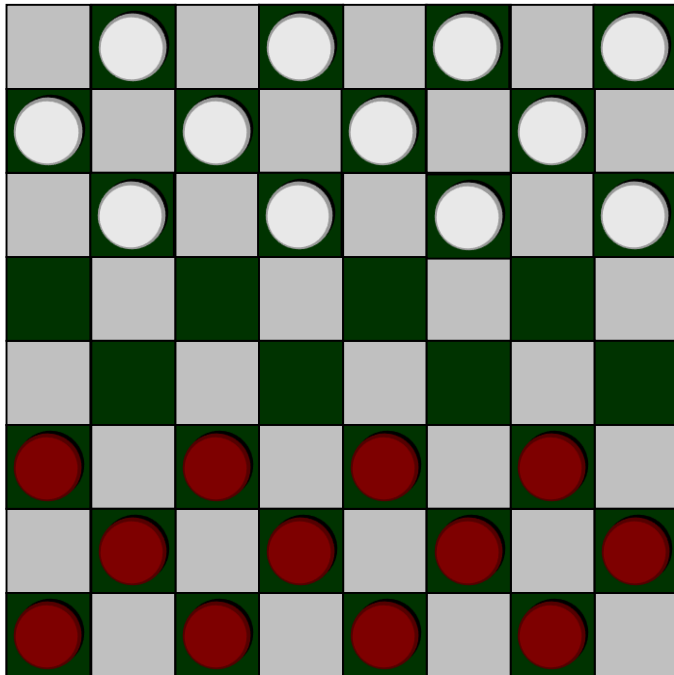
یادگیری ماشین: تعاریف

۱۰

□ آرتور ساموئل. [۱۹۵۹]

« یک حوزه مطالعاتی که به ماشین‌ها توانایی یادگیری می‌دهد، بدون این که نیاز باشد این ماشین‌ها به طور صریح برنامه‌نویسی شوند. »

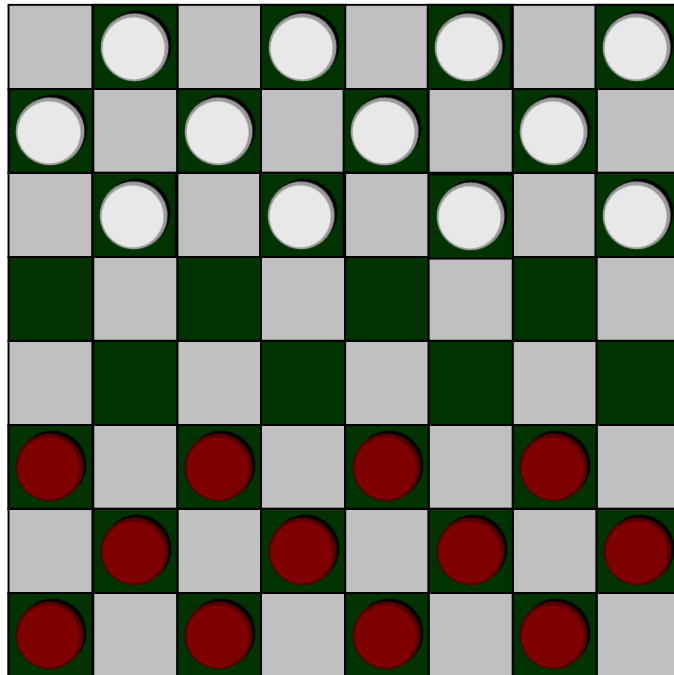
□ بازی چکرز. [ساموئل، دهه ۵۰]



یادگیری ماشین: تعاریف

□ تام میشل. [۱۹۹۸]

« با داشتن یک وظیفه مانند T و یک معیار کارایی مانند P، می‌گوییم یک برنامه کامپیوتری از تجربه‌ی E یاد می‌گیرد اگر معیار کارایی آن برنامه یعنی P برای انجام وظیفه T با استفاده از تجربه E بهبود یابد. »



□ مثال. بازی چکرز

□ **وظیفه:** انجام بازی چکرز

□ **تجربه:** هزاران هزار بار بازی در مقابل خود

□ **معیار کارایی:** تعداد دفعات برد در برابر رقبای جدید

مثال: تشخیص هرزنامه

۱۲

The screenshot shows a Gmail interface with a phishing email. The email is from **accounts-noreply@google.com** and is titled "Google Account recovery phone number changed". The sender's name is "accounts-noreply@google.com" and the recipient is "me". The email content includes a warning about a changed recovery phone number and links to support pages. The "Report spam" button is highlighted with a red box. The email content includes a warning about a changed recovery phone number and links to support pages.

Google Account recovery phone number changed

accounts-noreply@google.com to me 4:43 PM (0 minutes ago)

The recovery phone number for your Google Account - razavinaser@gmail.com - was recently changed. If you made this change, you don't need to do anything more.

If you didn't change your recovery phone, someone may have broken into your account. Visit this link for more information:
<https://support.google.com/accounts/bin/answer.py?answer=2450236>

If you are having problems accessing your account, begin the account recovery process by clicking this link:
<https://accounts.google.com/RecoverAccount?fpOnly=1&source=ancrpp&Email=razavinaser@gmail.com>

Important account security tips:

- Make sure you have regular access to your recovery phone so you can easily pass Google's security measures if you ever forget your password.
- Always keep your recovery phone number current. A phone is more secure than a recovery email or your security question because, unlike the other two, you usually have physical possession of your phone.

Note: This email address cannot accept replies.

Sincerely,
The Google Accounts Team

© 2012 Google Inc. 1600 Amphitheatre Parkway, Mountain View, CA 94043

You have received this mandatory email service announcement to update you about important changes to your Google product or account.

مثال: تشخیص هرزنامه

□ مثال. تشخیص هرزنامه

فرض کنید برنامه ایمیل شما به شما امکان می‌دهد که ایمیل‌های دریافتی خود را به عنوان هرزنامه علامت بزنید و بر این اساس یاد می‌گیرد که چگونه هرزنامه‌ها را بهتر فیلتر کند.

□ **وظیفه:** دسته‌بندی ایمیل‌ها به عنوان هرزنامه یا ایمیل.

□ **تجربه:** نظارت بر این که شما کدام ایمیل‌ها را به عنوان هرزنامه علامت می‌زنید.

□ **معیار کارایی:** تعداد ایمیل‌هایی که به درستی دسته‌بندی شده‌اند.

انواع روش‌های یادگیری ماشین

□ یادگیری ماشین. بهبود عملکرد ماشین در انجام یک وظیفه با کسب تجربه.

□ س. یک ماشین از کجا می‌تواند بفهمد عملکردش بهبود یافته است؟

□ می‌توانیم به ماشین پاسخ درست را برای چند نمونه محدود از ورودی‌ها بدهیم به این امید که بتواند آن را برای نمونه‌های دیگر تعمیم دهد -- **یادگیری نظارت شده**

□ می‌توانیم به ماشین بگوییم پاسخش تا چه میزان درست بوده (مثلا با دادن یک امتیاز) و خود ماشین مسئول یافتن پاسخ‌های درست است -- **یادگیری تقویتی**

□ ممکن است هیچ اطلاعاتی در مورد پاسخ درست به ماشین ندهیم و تنها از ماشین بخواهیم ورودی‌هایی را که دارای وجوه مشترک هستند پیدا کند -- **یادگیری بدون نظارت**

یادگیری نظارت شده

یادگیری نظارت شده

□ ورودی. یک مجموعه آموزشی که در آن به ازای هر ورودی پاسخ درست داده شده است.

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

مجموعه آموزشی

□ هدف. یافتن یک تقریب مناسب برای نگاشت زیر:

$$f: X \rightarrow Y$$

□ مثال.

□ تشخیص هرزنامه: نگاشت ایمیل‌ها به مجموعه {هرزنامه، غیرهرزنامه}

□ تشخیص ارقام: نگاشت یک مجموعه از پیکسل‌ها به مجموعه {۰, ۱, ۲, ..., ۹}

□ تشخیص سرطان: نگاشت داده‌های پزشکی به مجموعه {بدخیم، خوش‌خیم}

مثال: تشخیص هرزنامه

۱۷

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret. ...



To be removed from future mailings, simply reply to this message and put "remove" in the subject.

99 million email addresses for only \$99



Ok, I know this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.



ورودی. ایمیل

خروجی. هرزنامه، غیرهرزنامه

مثال: تشخیص ارقام دست‌نویس

8	3	9	3	8	5	8	5	6	5
9	4	9	5	7	1	7	6	1	1
6	8	3	6	8	8	8	1	1	4
4	9	5	0	1	2	1	4	5	3
7	2	7	2	6	3	1	1	2	1
3	2	7	0	4	6	0	8	1	8
6	0	7	4	1	1	7	4	2	1
2	9	5	3	7	4	1	0	5	8
3	5	5	7	6	5	9	9	9	3
1	9	9	6	1	2	1	3	6	7

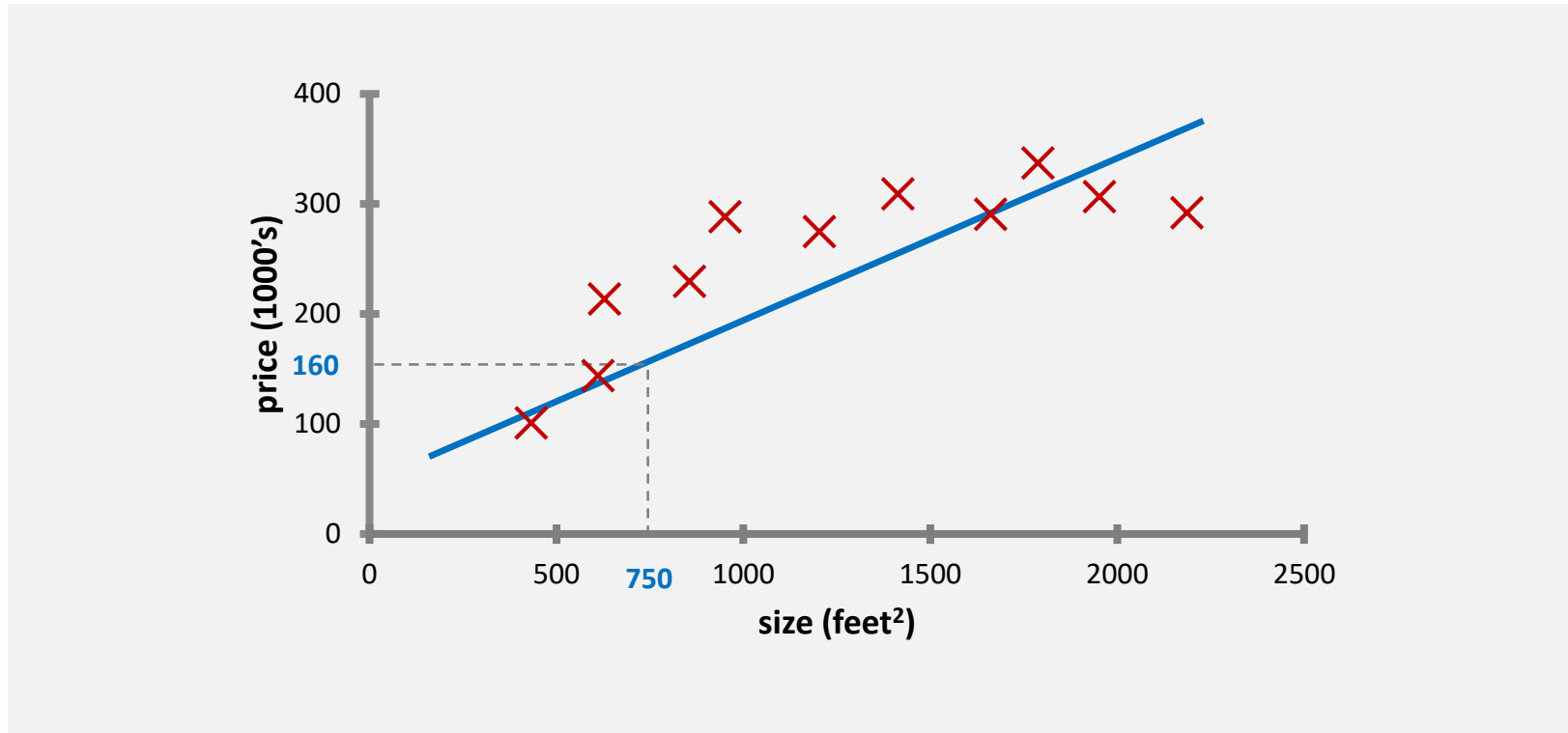
- ورودی. تصویر یک رقم
- خروجی. یک رقم

مثال: قیمت‌گذاری یک خانه

۱۹

□ ورودی. اندازه خانه [برحسب فوت مربع]

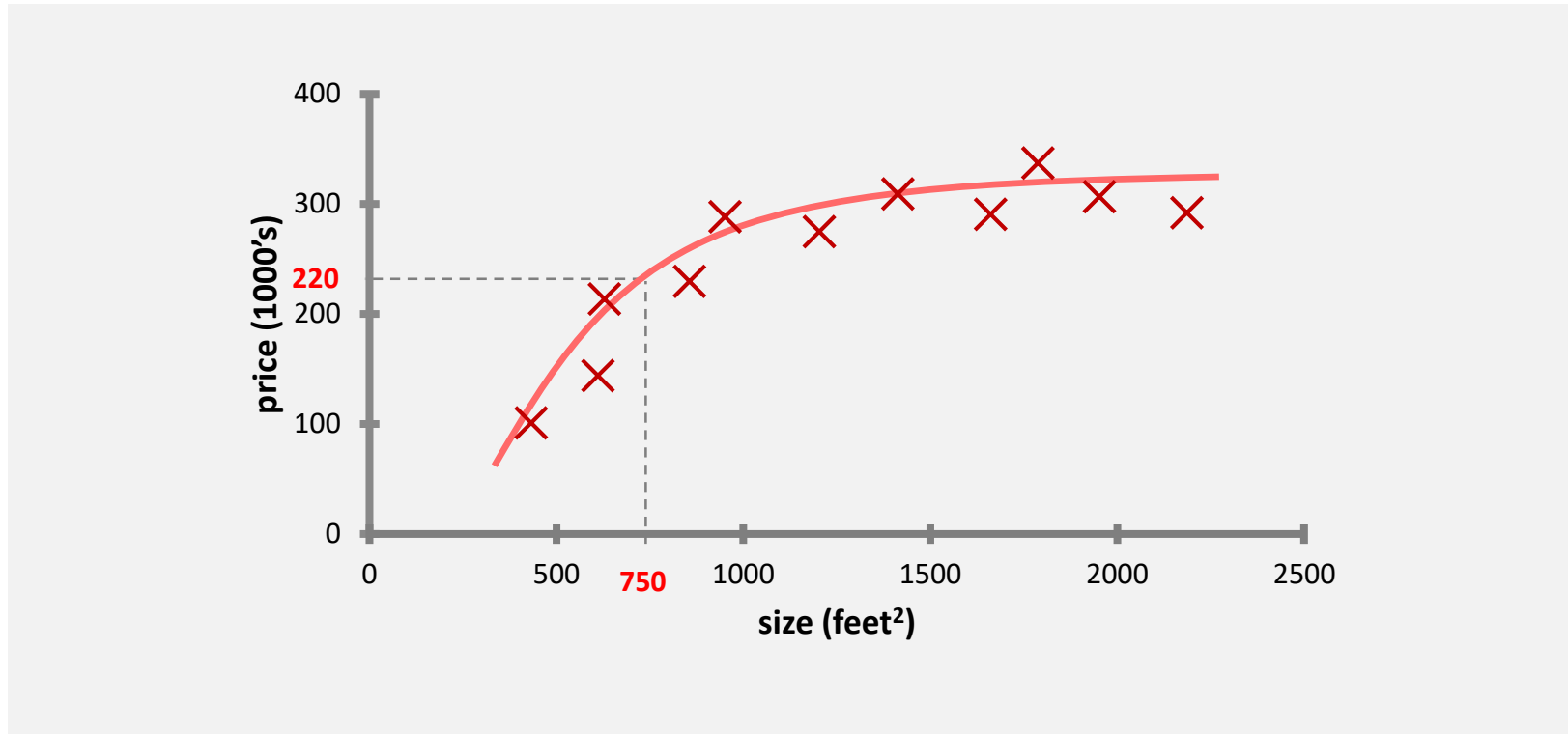
□ خروجی. قیمت تخمینی



مثال: قیمت‌گذاری یک خانه

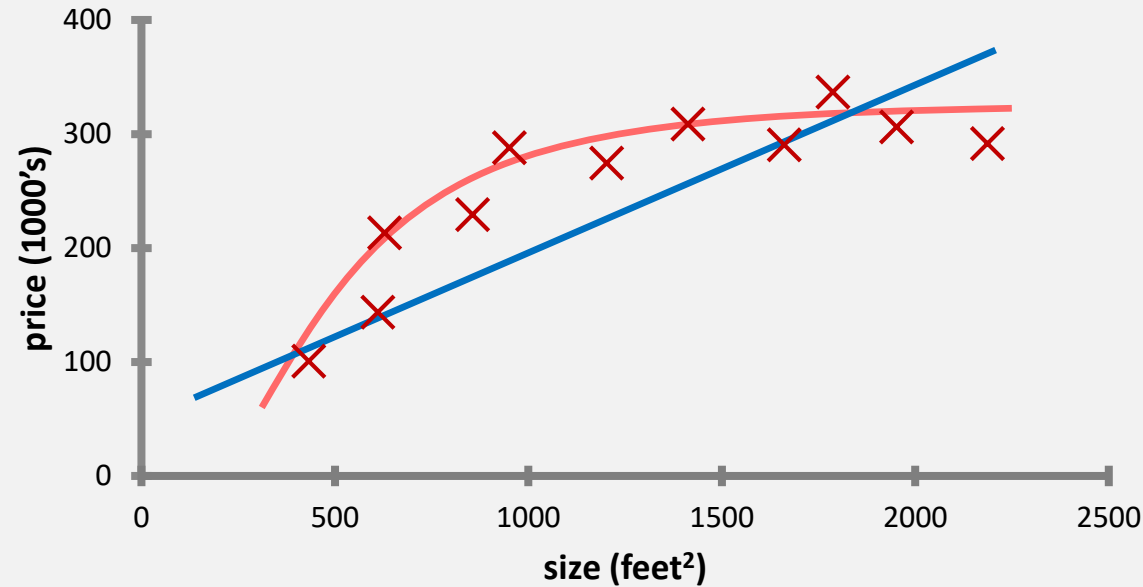
۲۰

□ س. کدام یک بهتر است؟ یک تابع خطی یا یک تابع درجه دوم؟



مثال: قیمت‌گذاری یک خانه

۲۱



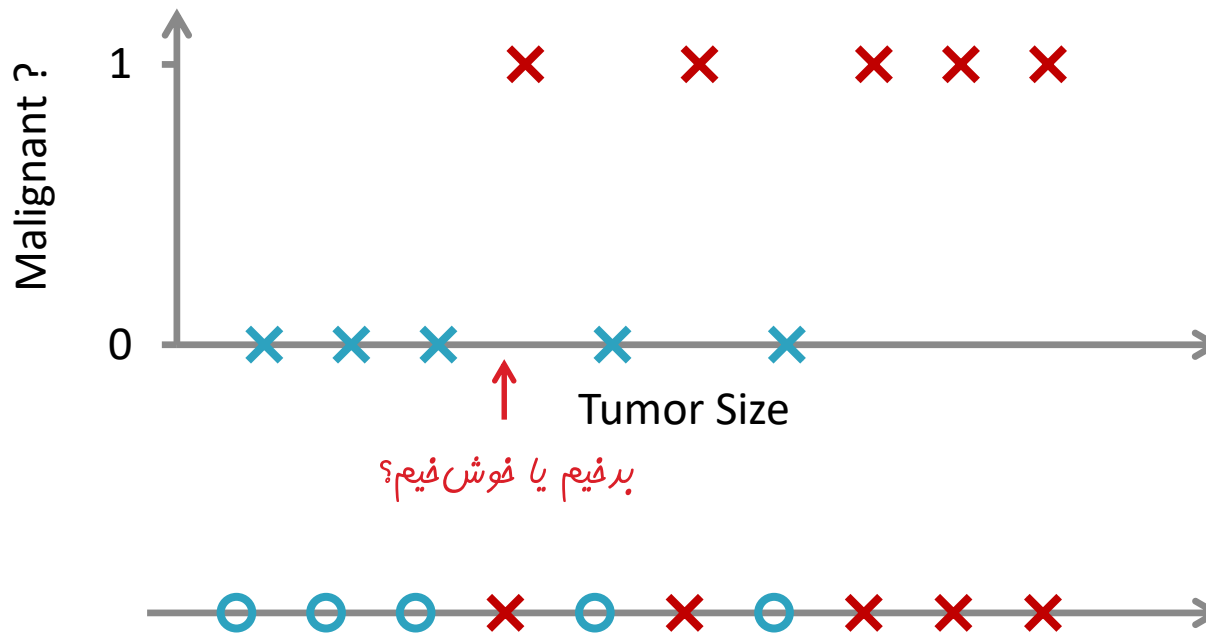
□ رگرسیون.

پیش‌بینی کمیت‌هایی با مقادیر پیوسته (مانند قیمت یک خانه)

□ یادگیری نظارت شده.

به ازای هر نمونه آموزشی، «پاسخ درست» داده شده است.

مثال: تشخیص نوع سرطان (بدخیم، خوش خیم)



□ کلاس بندی.

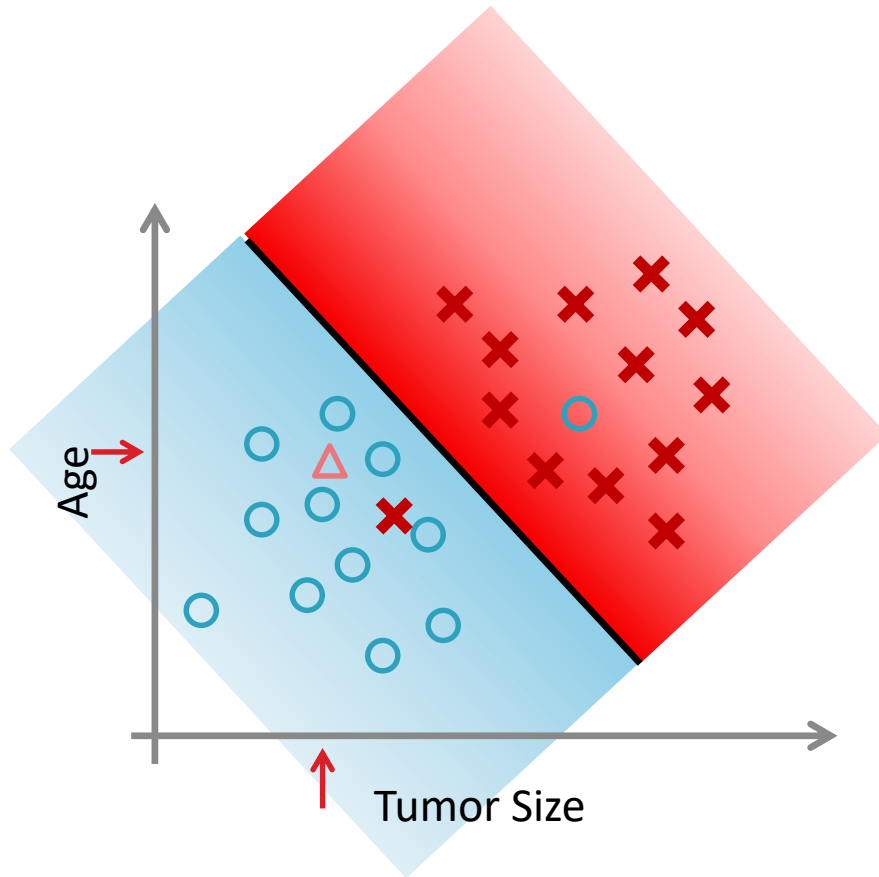
پیش بینی کمیت‌هایی با مقادیر **گسسته** (مانند صفر و یک).

□ یادگیری نظارت شده.

به ازای هر نمونه آموزشی، «پاسخ درست» داده شده است.

مثال: تشخیص نوع سرطان (بدخیم، خوش خیم)

۲۳



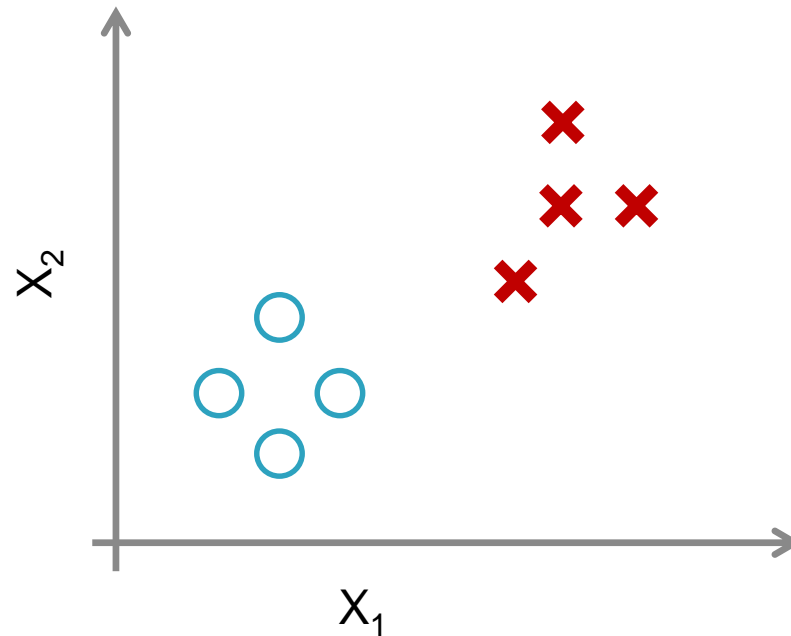
□ ویژگی‌های دیگر.

- یکنواختی اندازه سلول‌ها
- یکنواختی شکل سلول‌ها
- ...

یادگیری بدون نظارت

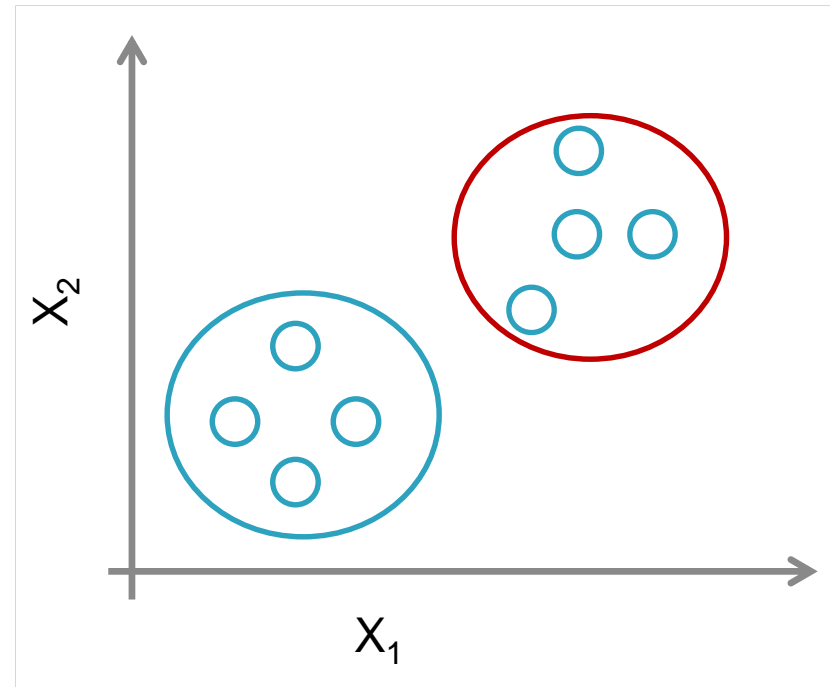
یادگیری نظارت شده

□ یادگیری نظارت شده. به ازای هر نمونه، پاسخ درست داده شده است.



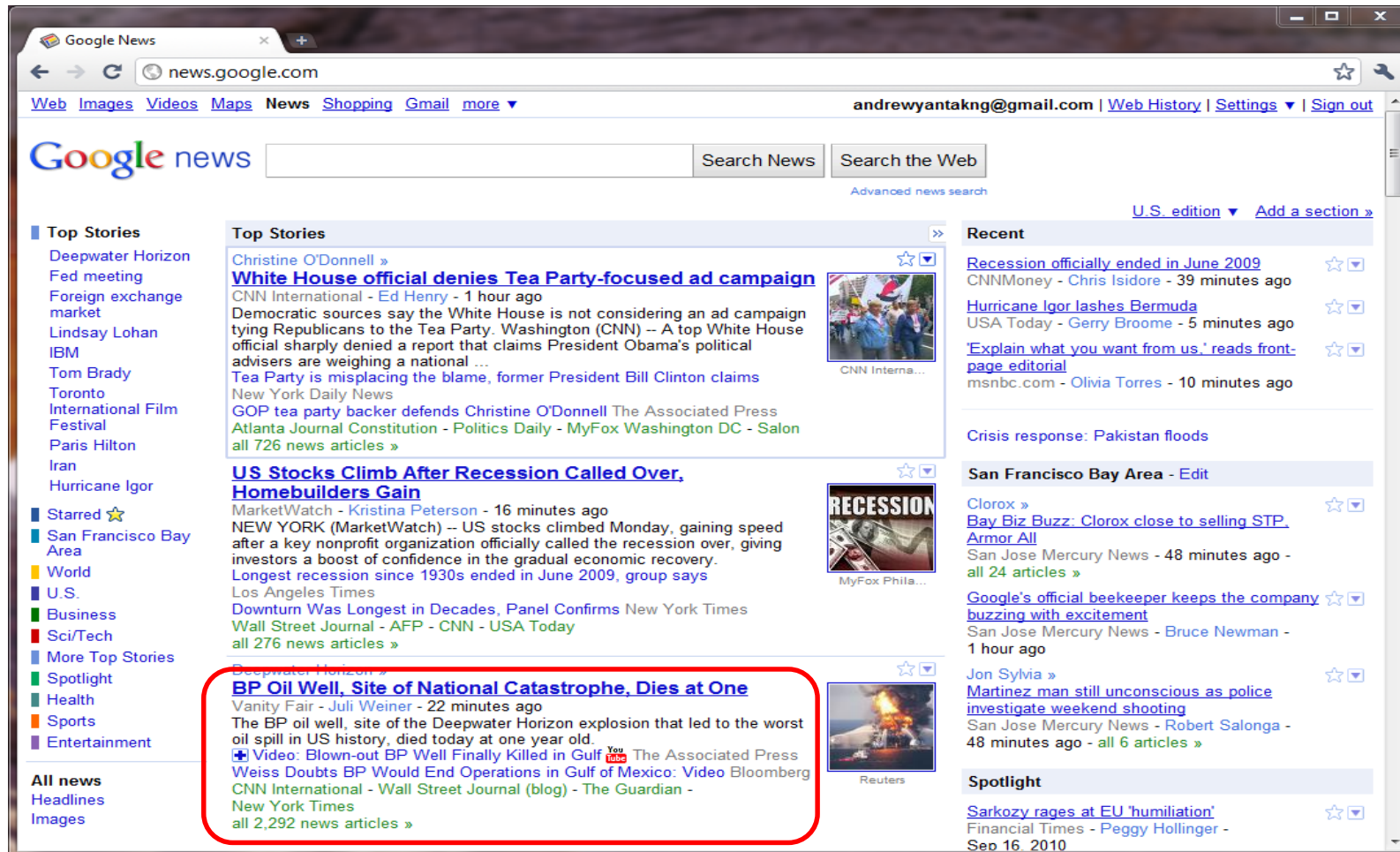
یادگیری بدون نظارت

□ یادگیری بدون نظارت. هیچ گونه اطلاعاتی در مورد پاسخهای درست داده نشده است!

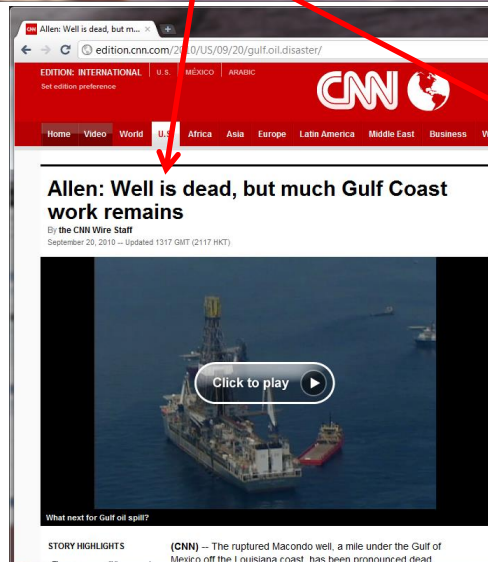
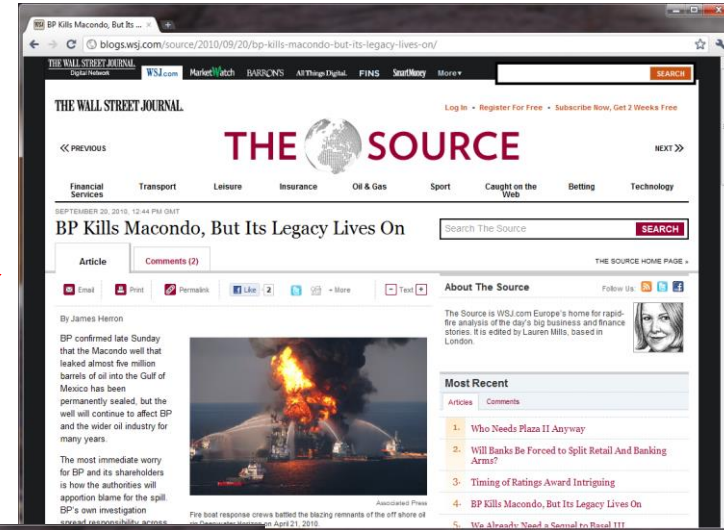
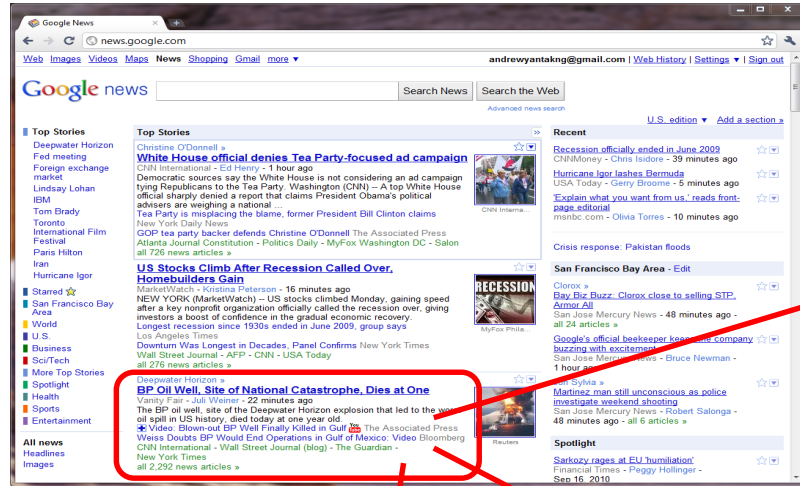


□ هدف. تشخیص ساختار در داده‌های ورودی (گروه‌بندی داده‌های مشابه).

کاربرد خوشه‌بندی: گروه‌بندی اخبار مرتباً



کاربرد خوشه‌بندی: گروه‌بندی اخبار مرتبط

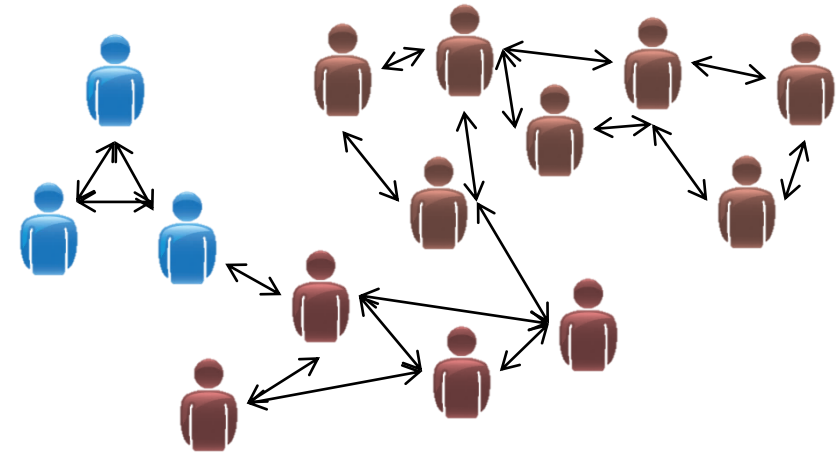


چند کاربرد دیگر از یادگیری بدون نظارت

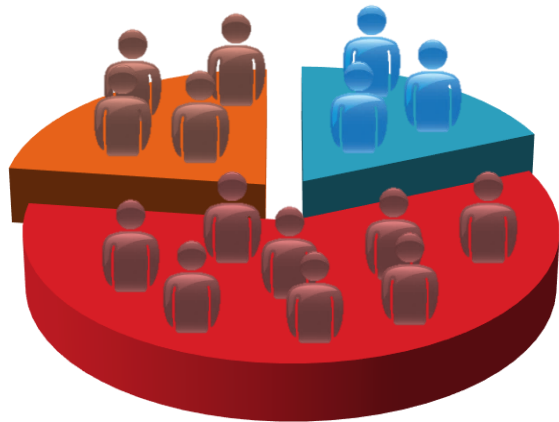
۲۹



سازمان‌دهی کلاسترهای محاسباتی (مرکز داده‌ها)



تحلیل شبکه‌های اجتماعی



بفش‌بندی بازار

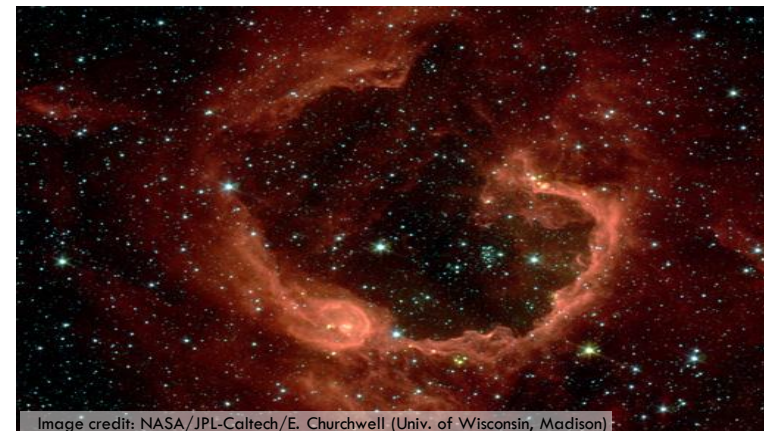
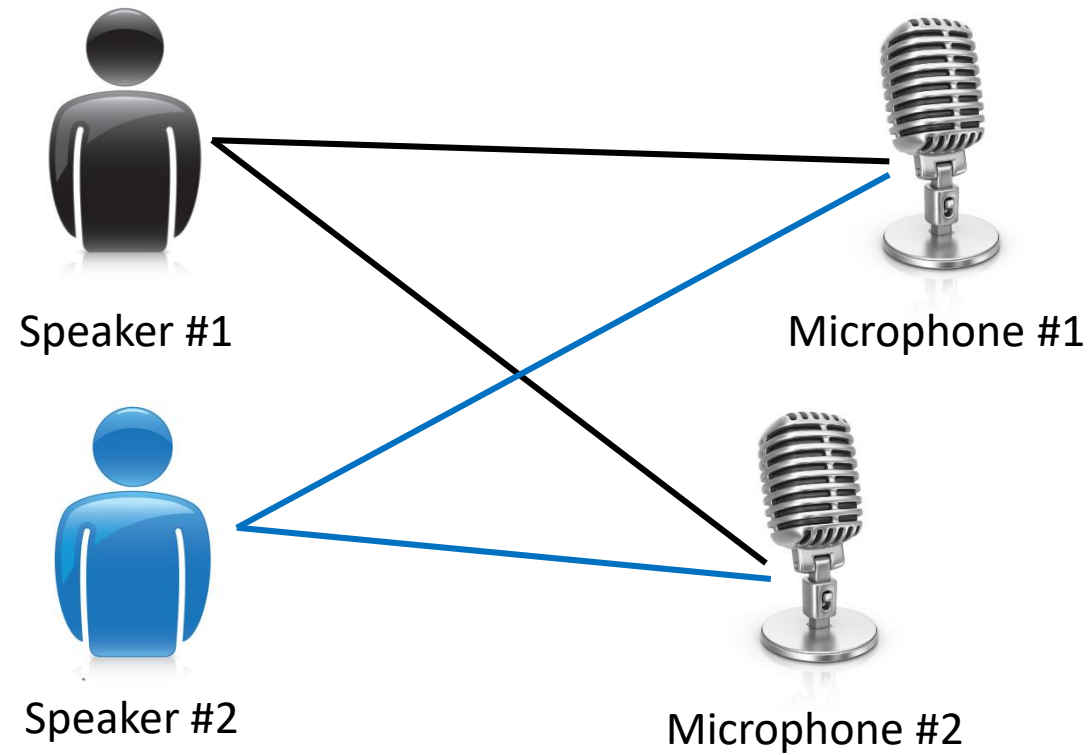


Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, Madison)

تحلیل داده‌های ستاره‌شناسی (نوعه تشکیل کوهکشان‌ها)

مسئله جشن کوکتل

۳۰



مسئله جشن کوکتل

۳۱

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️

الگوریتم مسئله جیشن کوکتل

۳۲

□ کد اکتاو.

```
[W, s, v] = svd(( repmat(sum(x .* x, 1), size(x, 1), 1) .* x) * x');
```


□ برای کدام یک از مسائل داده شده زیر باید از یک الگوریتم یادگیری بدون نظارت استفاده شود؟

- توسعه یک برنامه برای فیلتر کردن هرزنامه‌ها با داشتن تعدادی ایمیل معمولی و تعدادی هرزنامه
- گروه‌بندی یک مجموعه از مقالات جدید یافته شده در وب بر اساس موضوع
- گروه‌بندی مجموعه‌ای از مشتری‌ها در چند بخش مختلف بازار با داشتن یک پایگاه داده در مورد مشتریان
- تشخیص دیابت در بیماران جدید با داشتن داده‌های مربوط به تعدادی فرد سالم و دیابتی