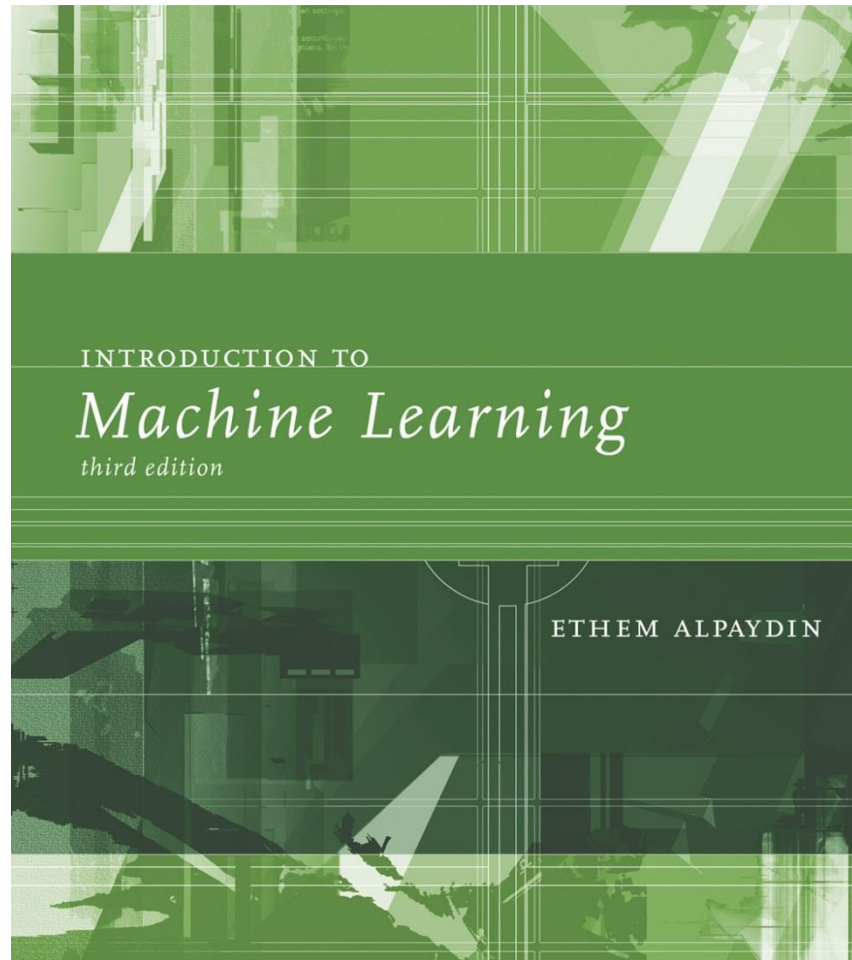


# یادگیری ماشین: معرفی

سید ناصر رضوی [n.razavi@tabrizu.ac.ir](mailto:n.razavi@tabrizu.ac.ir)

۱۳۹۵

# منابع و مراجع (۱)



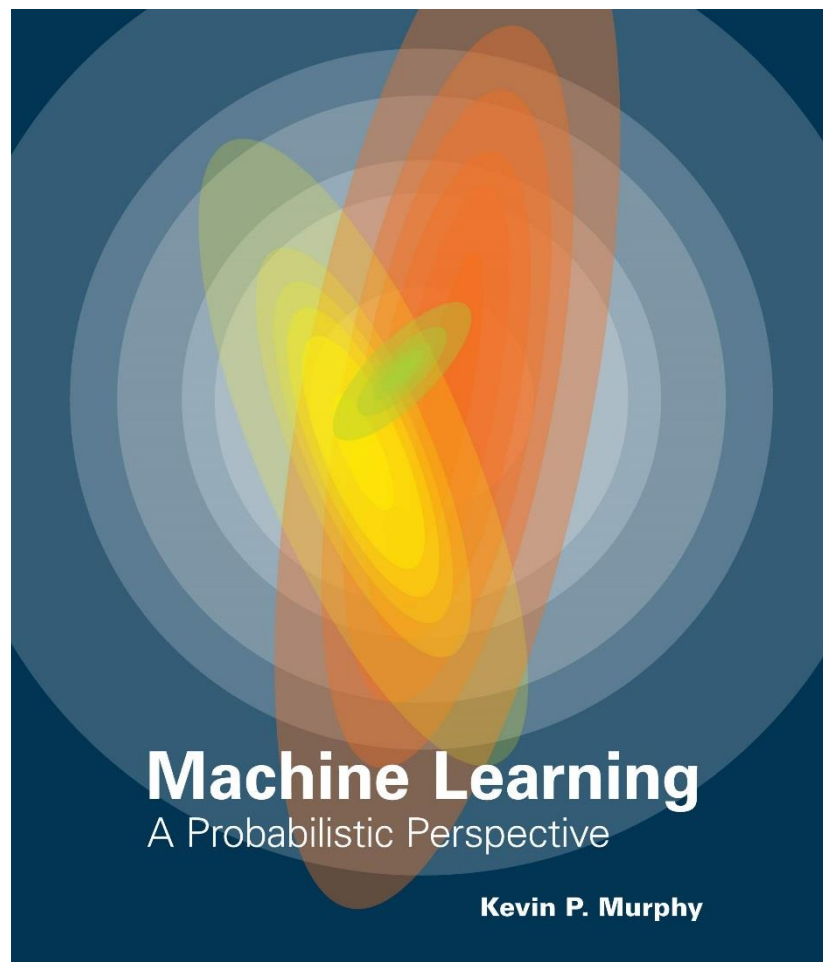
□ مقدمه‌ای بر یادگیری ماشین.  
[آلپایدین، ویراست سوم؛ ۲۰۱۴]

□ یادگیری ماشین: یک دیدگاه احتمالاتی.  
[کونین مورفی، ۲۰۱۲]

□ شناسایی الگو و یادگیری ماشین.  
[کریستوفر بیشاپ، ۲۰۰۶]

## منابع و مراجع (۲)

۳



□ مقدمه‌ای بر یادگیری ماشین.

[آلپایدین، ویراست سوم؛ ۲۰۱۴]

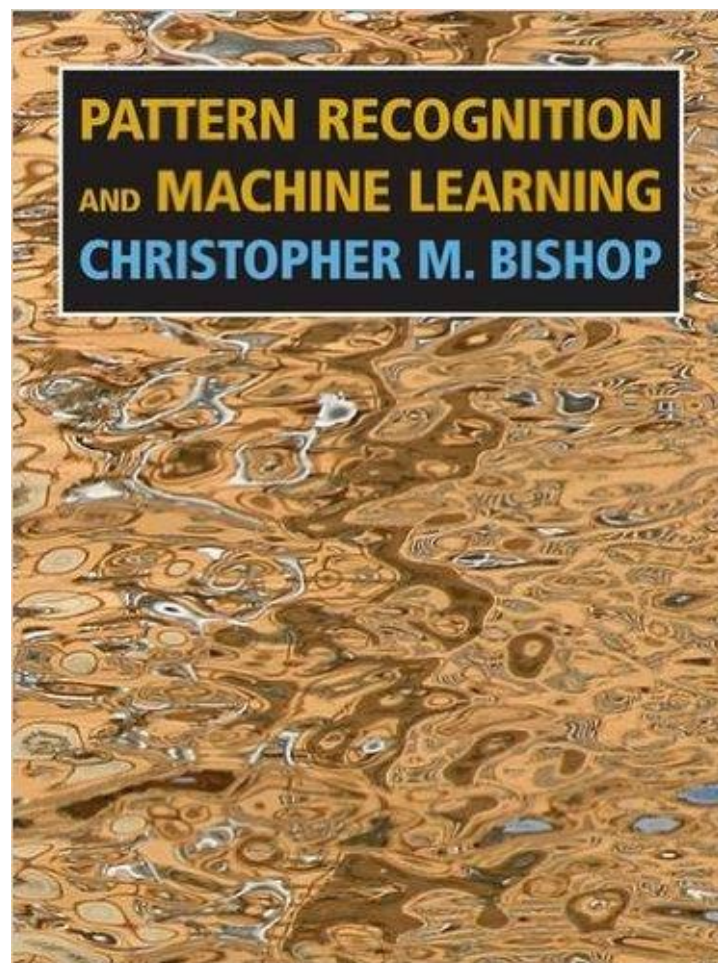
□ یادگیری ماشین: یک دیدگاه احتمالاتی.

[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]

# پیش‌نیازها

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

## □ جبر خطی

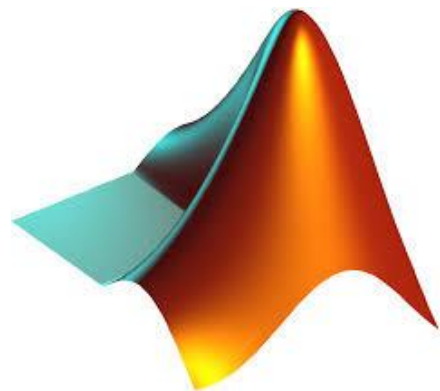
- ماتریس‌ها، بردارها، عملیات ماتریسی و دستگاه معادلات خطی
- ماتریس وارون، بردارهای ویژه، مرتبه ماتریس، تجزیه مقادیر منفرد

## □ حساب چند متغیره

- مشتق، انتگرال، صفحات مماس

## □ احتمالات

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



## MATLAB



## OCTAVE

□ تمرین‌ها [۴۰٪]

□ مباحث نظری

□ برنامه‌نویسی

□ امتحان پایان‌ترم [۵۰٪]

□ حضور موثر در کلاس [۱۰٪]

# 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 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



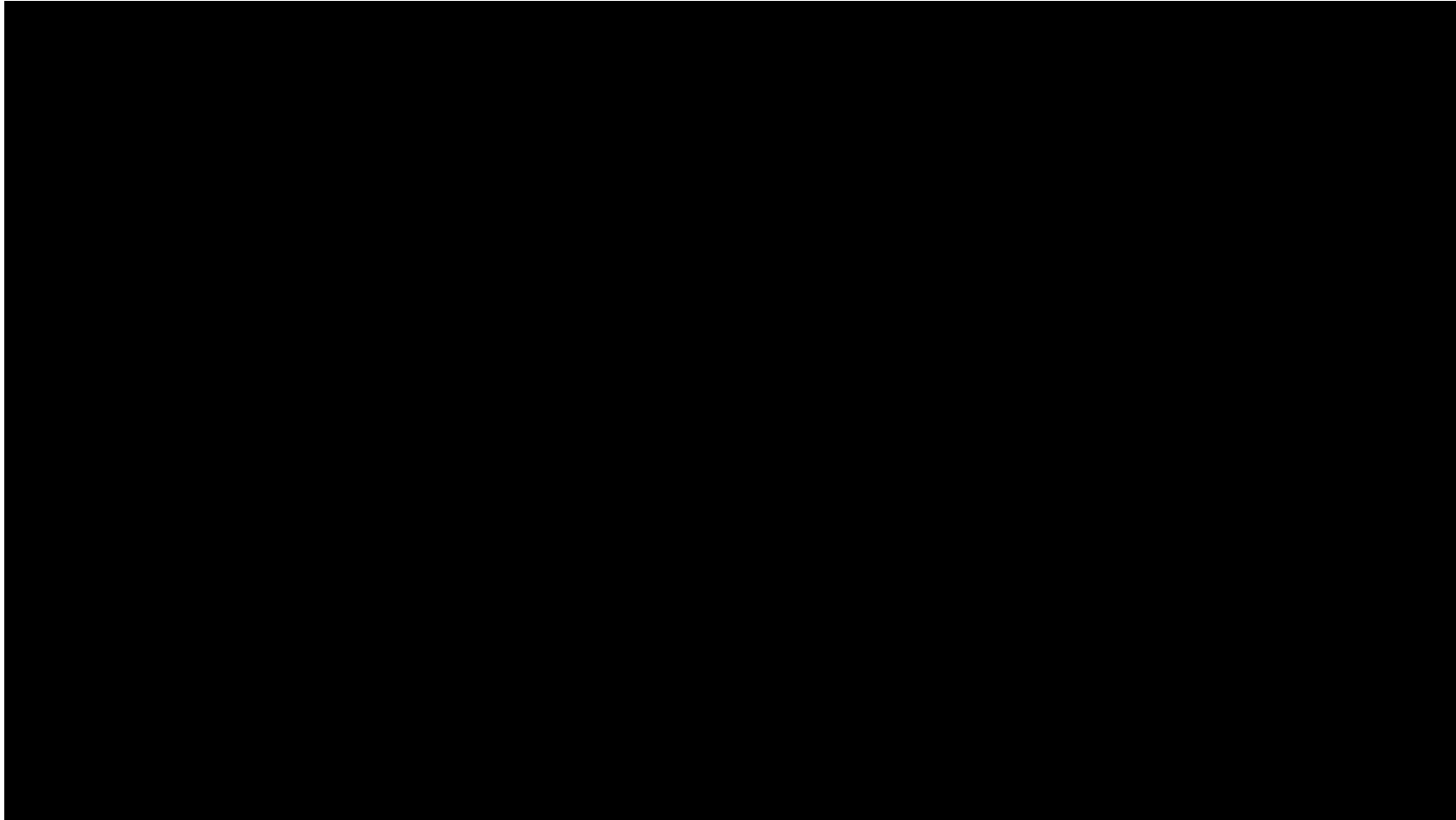
«یادگیری ماشین در نهایت به یک انقلاب واقعی منجر خواهد شد»

گرگ پاپادوپولس - مدیر اسبق سان





# مثال: بالگرد بدون خلبان



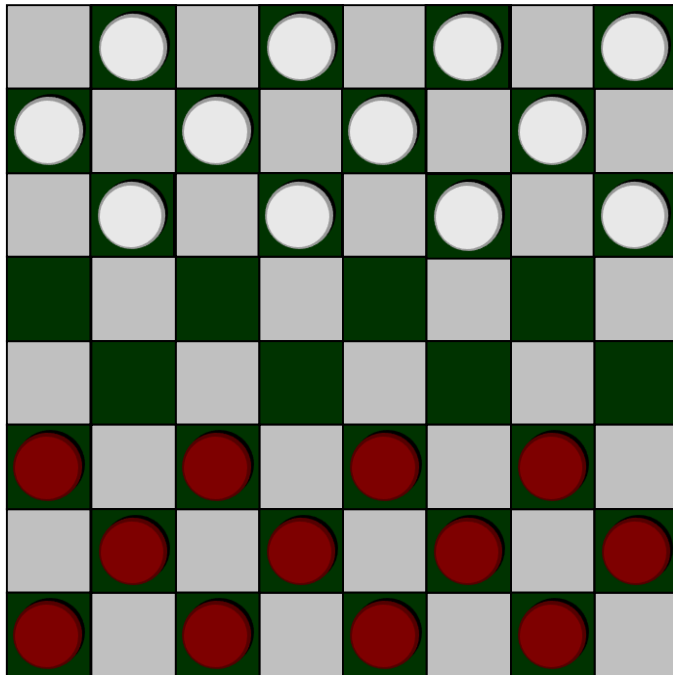
# یادگیری ماشین چیست؟

# یادگیری ماشین: تعاریف

□ آرتور ساموئل. [۱۹۵۹]

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

□ بازی چکرز. [ساموئل، دهه‌ی ۵۰]

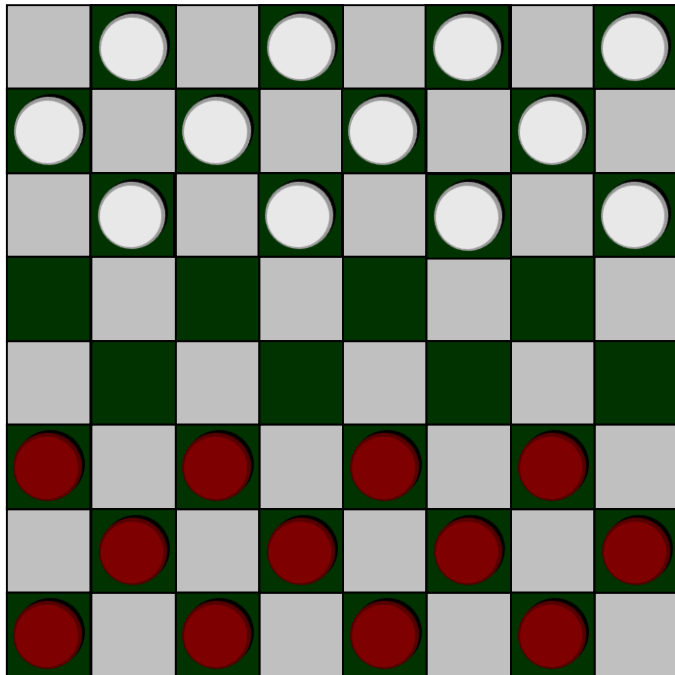


# یادگیری ماشین: تعاریف

۱۲

□ تام میشل. [۱۹۹۸]

« با داشتن یک وظیفه مانند T و یک معیار کارایی مانند P، می‌گوییم یک برنامه‌ی کامپیوتری از تجربه‌ی E یاد می‌گیرد اگر معیار کارایی آن برنامه یعنی P برای انجام وظیفه‌ی T با استفاده از تجربه‌ی 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

۱۳

https://mail.google.com/mail/?shva=1#inbox/13c008bb49a5ac10

+You Search Images Maps Play YouTube News Gmail Drive Calendar More

Google

razavinaser@gmail.com

New! Gmail's mobile apps just got updated on Google Play and the Apple App Store. Dismiss

Report spam

COMPOSE

Inbox (7)

Starred

Important

Chats

Sent Mail

Drafts

All Mail

Spam (17)

Trash

Search people...

razavi

Google Account recovery phone number changed

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

The recovery phone number for your Google Account - [razavinaser@gmail.com](mailto: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.

Ads - Why these ads?

**Email Auto-Responders**  
Quickly Engage New Leads & Contacts Email Auto-Responder. Free Account!  
[www.Contactology.com/AutoRespon](http://www.Contactology.com/AutoRespon)

**Do You Carry Concealed?**  
Know Your Rights & Get Your Free Concealed Carry Report Today!  
[USConcealedCarry.net](http://USConcealedCarry.net)

**Try GFI® VIPRE Free**  
Award-winning antivirus software. Download a free 30-day trial now.  
[www.VIPREbusiness.com](http://www.VIPREbusiness.com)

**2013 Grants**  
Grant Funding May Be Available See If You Qualify!  
[www.ClassesUSA.com](http://www.ClassesUSA.com)

More about...

- [Password Recovery »](#)
- [Password Unlock »](#)
- [Password Security »](#)
- [Password Change »](#)

# مثال: تشخیص هرزنامه

## □ مثال. تشخیص هرزنامه

فرض کنید برنامه‌ی 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 نمونه‌ی 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 مجموعه‌ی آموزشی که در آن به ازای هر ورودی پاسخ درست داده شده است.

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

مجموعه‌ی آموزشی

□ هدف. یافتن یک تقریب مناسب برای نگاشت زیر:

$$f: X \rightarrow Y$$

□ مثال.

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

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

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

# مثال: تشخیص هرزنامه

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

۱۹

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

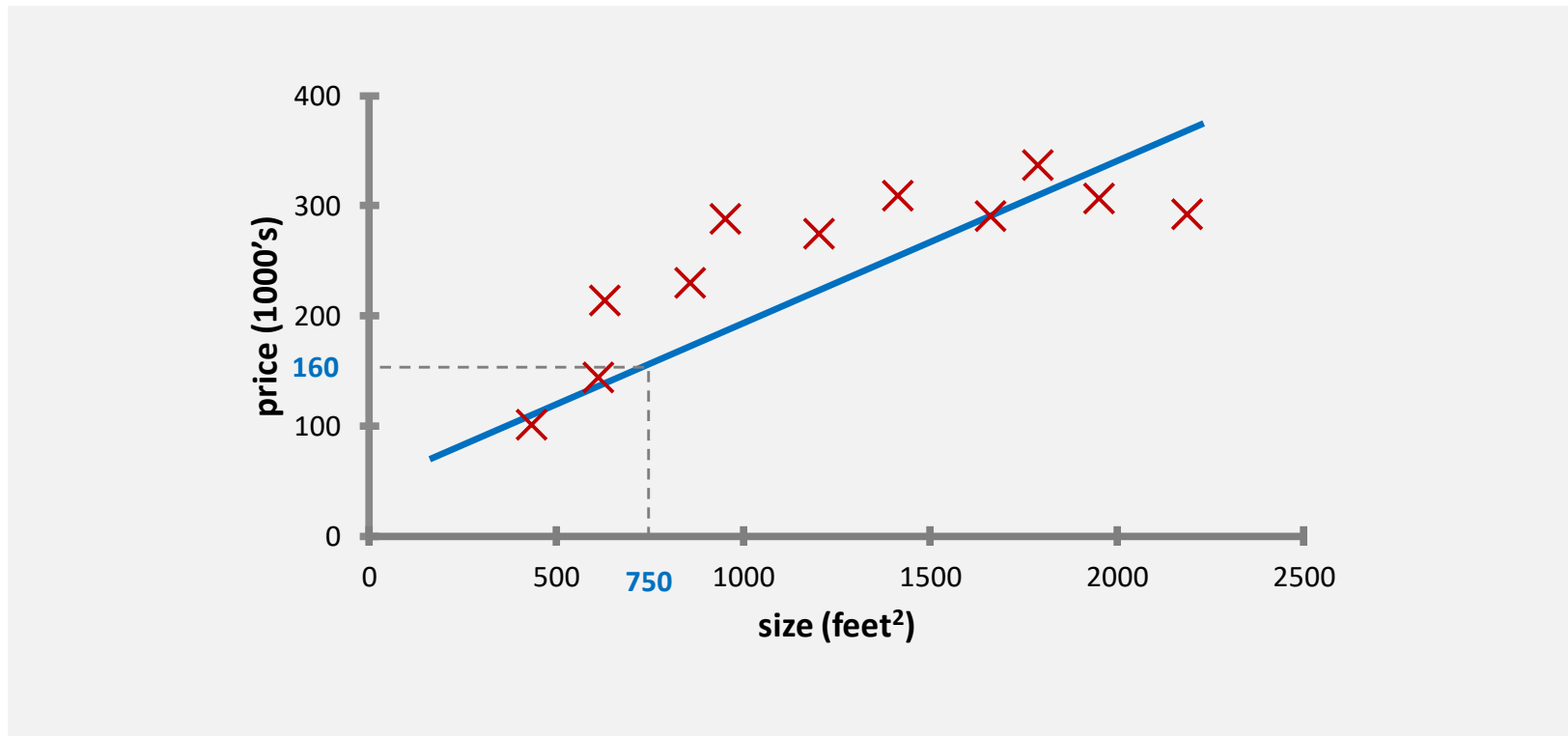
□ ورودی. تصویر یک رقم

□ خروجی. یک رقم

# مثال: قیمت‌گذاری یک خانه

۲۰

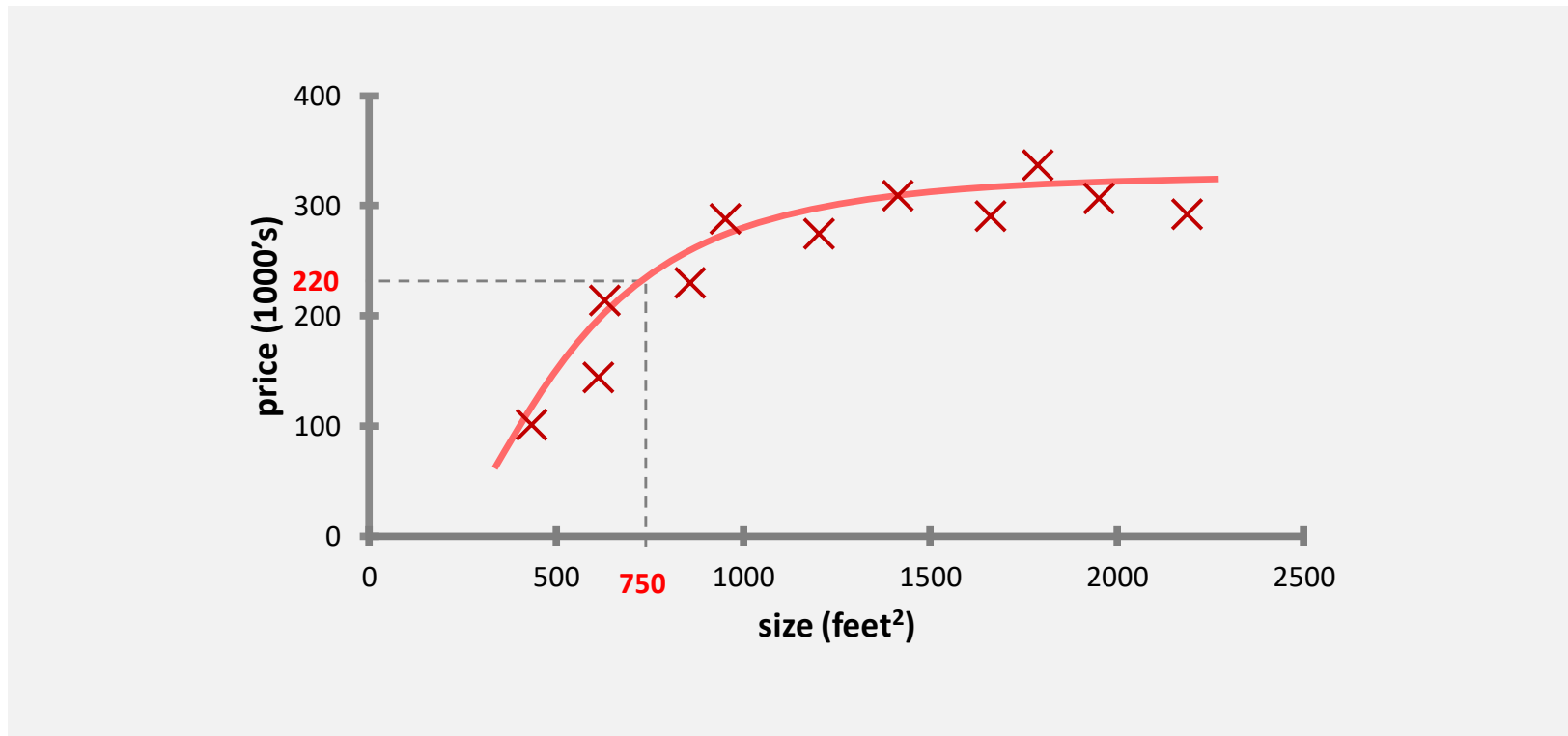
- ورودی. اندازه‌ی خانه [برحسب فوت مربع]
- خروجی. قیمت تخمینی



# مثال: قیمت‌گذاری یک خانه

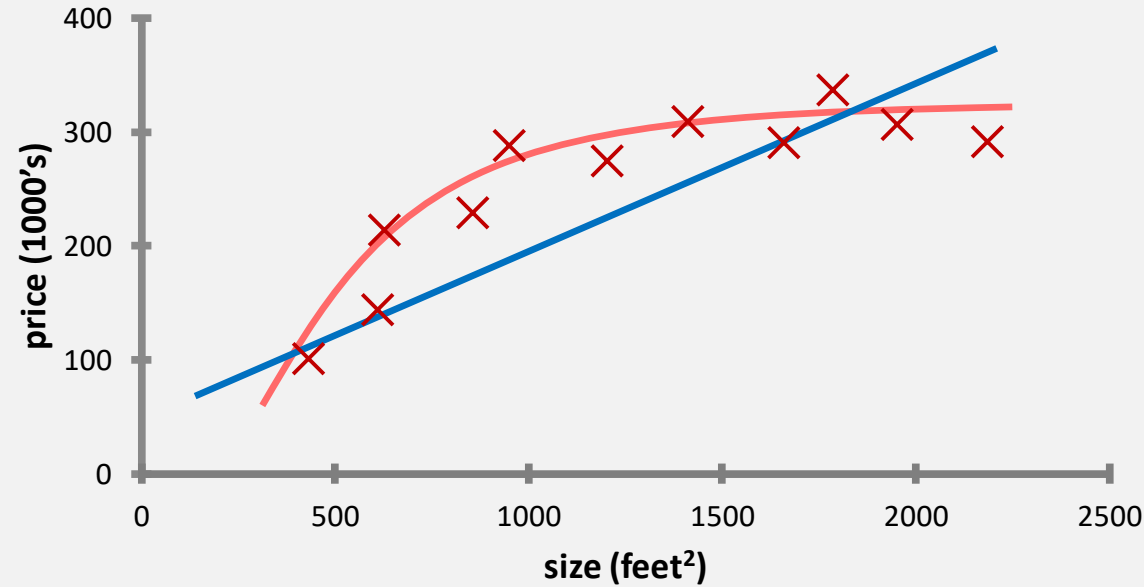
۲۱

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



# مثال: قیمت‌گذاری یک خانه

۲۲



□ رگرسیون.

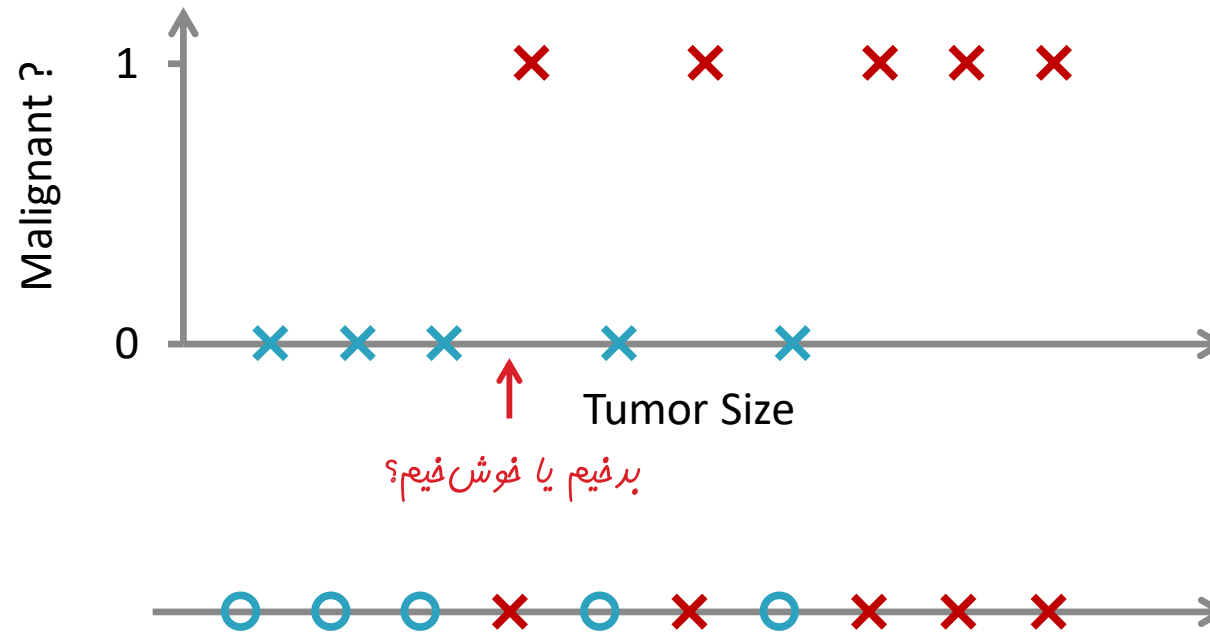
پیش‌بینی کمیت‌هایی با مقادیر پیوسته (مانند قیمت یک خانه)

□ یادگیری نظارت شده.

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

# مثال: تشخیص نوع سرطان (بدخیم، خوش خیم)

۲۳



□ کلاس بندی.

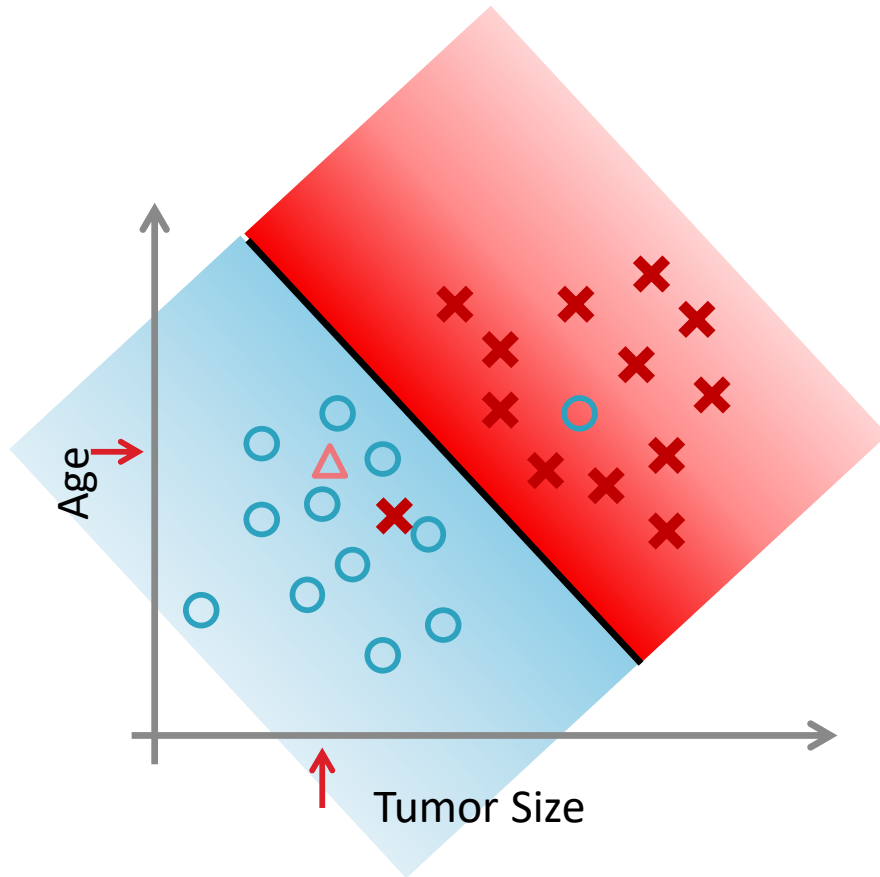
پیش بینی کمیت‌هایی با مقادیر **گسسته** (مانند صفر و یک).

□ یادگیری نظارت شده.

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

# مثال: تشخیص نوع سرطان (بدخیم، خوش خیم)

۲۴



□ ویژگی‌های دیگر.

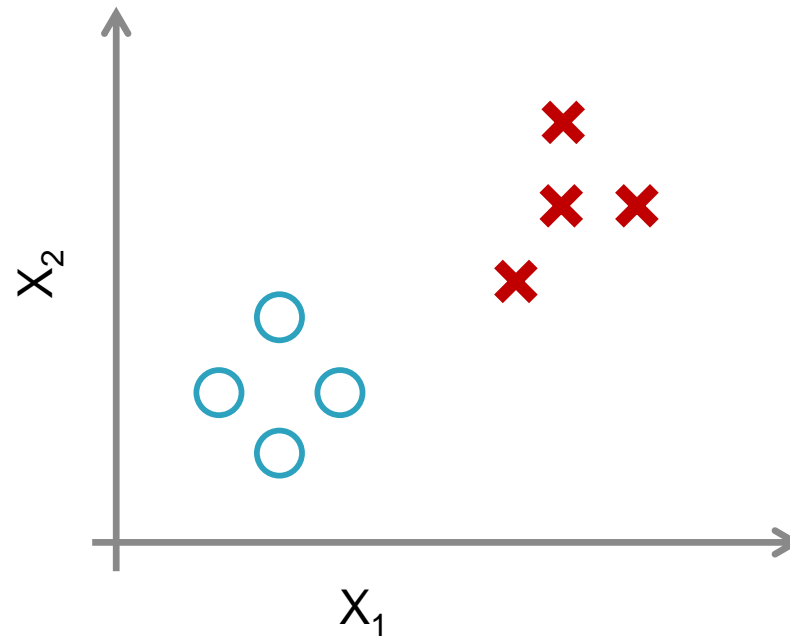
- یکنواختی اندازه سلول‌ها
- یکنواختی شکل سلول‌ها
- ...



# یادگیری بدون نظارت

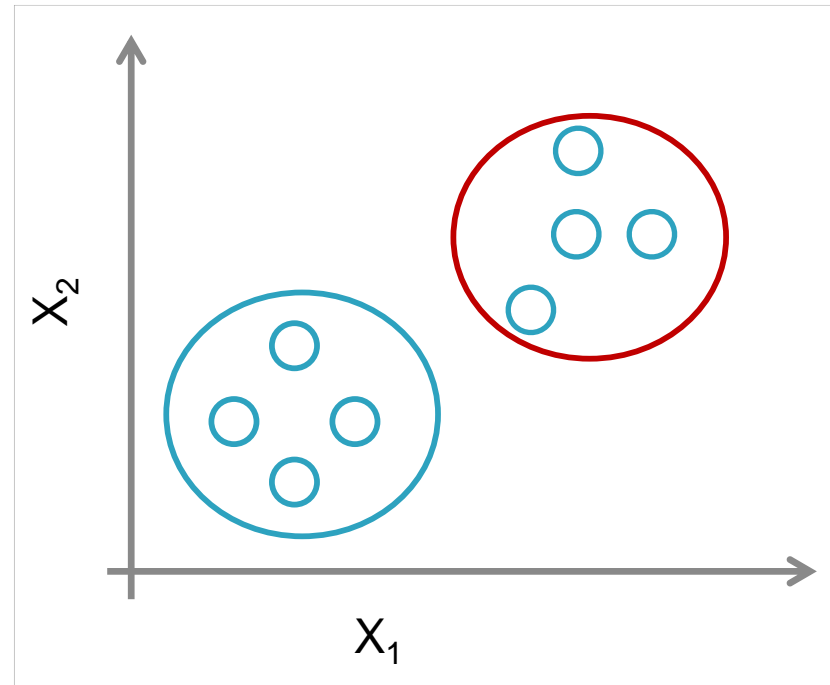
# یادگیری نظارت شده

□ 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!



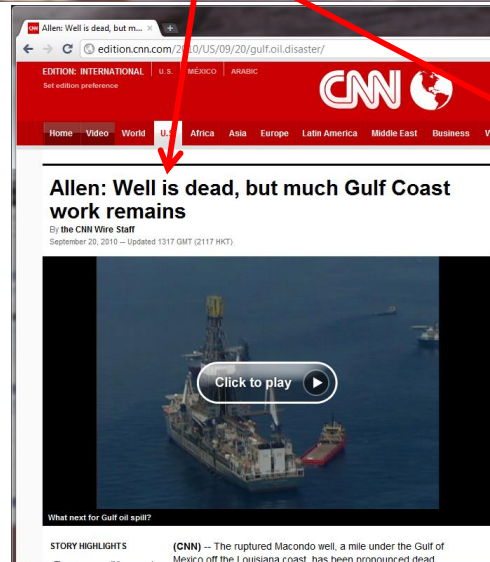
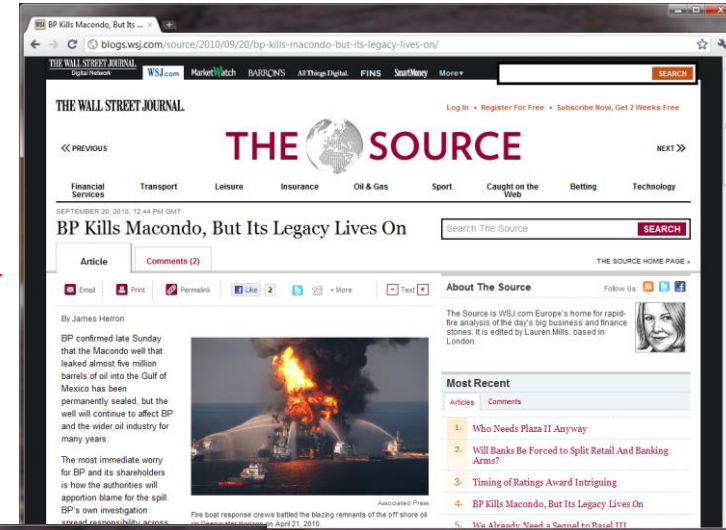
□ هدف. تشخیص ساختار در داده‌های ورودی (گروه‌بندی داده‌های مشابه).

# کاربرد خوشه‌بندی: گروه‌بندی اخبار مرتباً

The screenshot shows the Google News homepage in a browser window. The address bar displays 'news.google.com'. The page features a search bar, navigation links for 'Web', 'Images', 'Videos', 'Maps', 'News', 'Shopping', and 'Gmail'. The main content area is divided into several sections:

- Top Stories:** Includes articles such as 'White House official denies Tea Party-focused ad campaign' and 'US Stocks Climb After Recession Called Over, Homebuilders Gain'. A red box highlights the article 'BP Oil Well, Site of National Catastrophe, Dies at One'.
- Recent:** Lists recent news items like 'Recession officially ended in June 2009' and 'Hurricane Igor lashes Bermuda'.
- Spotlight:** Features a featured article 'Sarkozy rages at EU 'humiliation''.
- Left Sidebar:** Contains 'Top Stories' and 'Starred' sections with various news topics and categories.

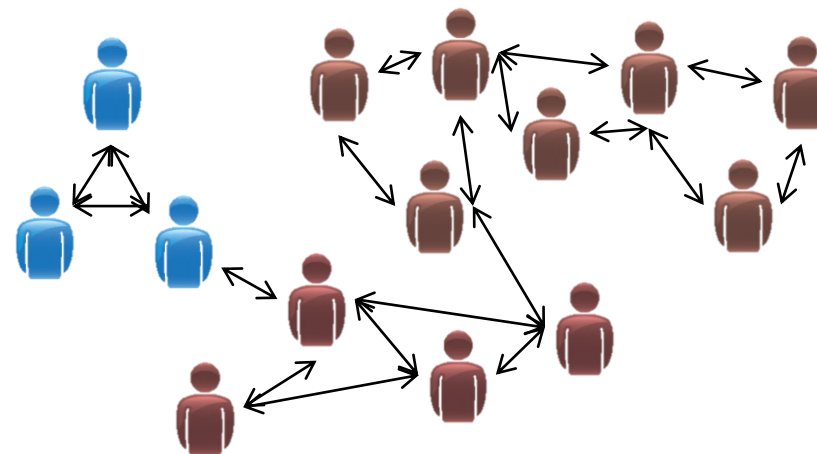
# کاربرد خوشه‌بندی: گروه‌بندی اخبار مرتباً



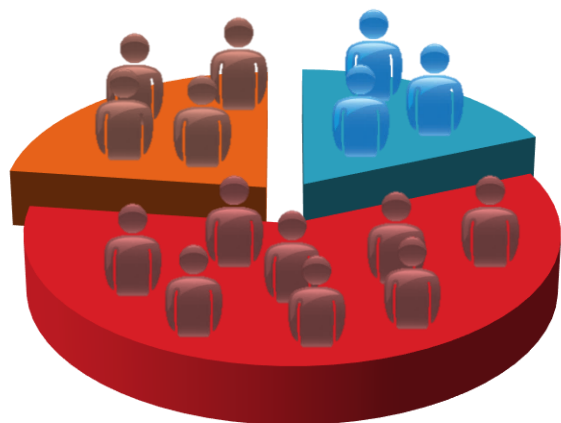
# چند کاربرد دیگر از یادگیری بدون نظارت



سازمان‌دهی کلاسترهای محاسباتی (مرکز داده‌ها)



تفیل شبکه‌های اجتماعی



بفش‌بندی بازار

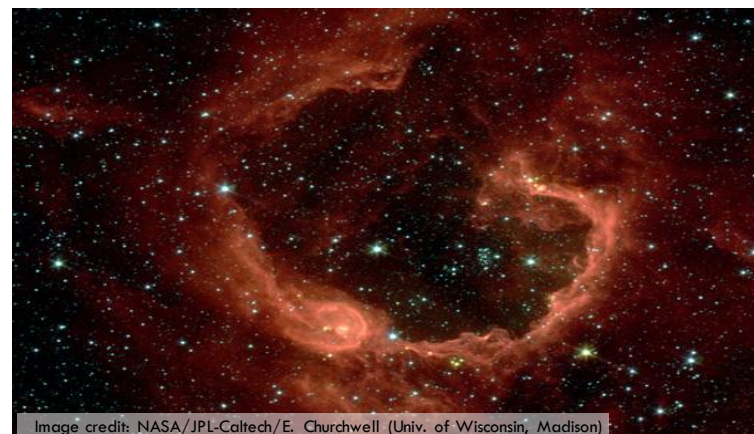
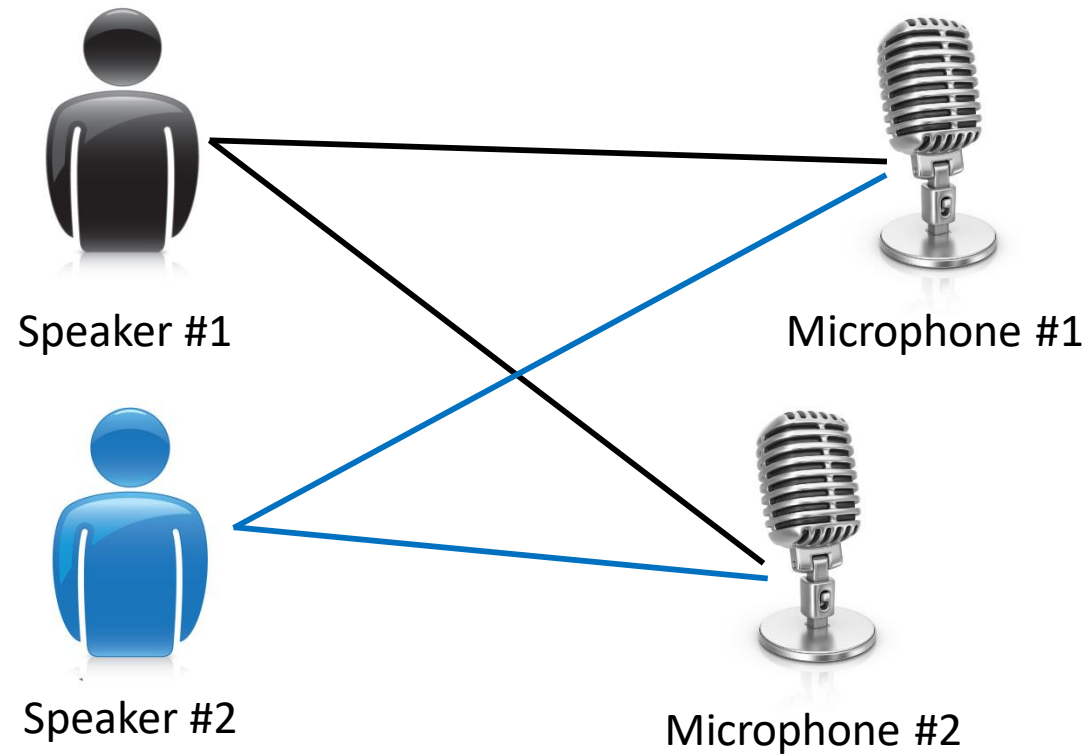


Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, Madison)

تفیل داده‌های ستاره‌شناسی (نوعه تشکیل کولکشان‌ها)

# مسئله‌ی جشن کوکتل

۳۱



# مسئله‌ی جشن کوکتل

۳۲

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️

---

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️



# الگوریتم مسئله‌ی جستن کوکتل

۳۳

□ 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؟

- توسعه‌ی 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