

# پردازش گفتار

## مرواری بر یادگیری ماشین

هادی ویسی

[h.veisi@ut.ac.ir](mailto:h.veisi@ut.ac.ir)

دانشگاه تهران - دانشکده علوم و فنون نوین

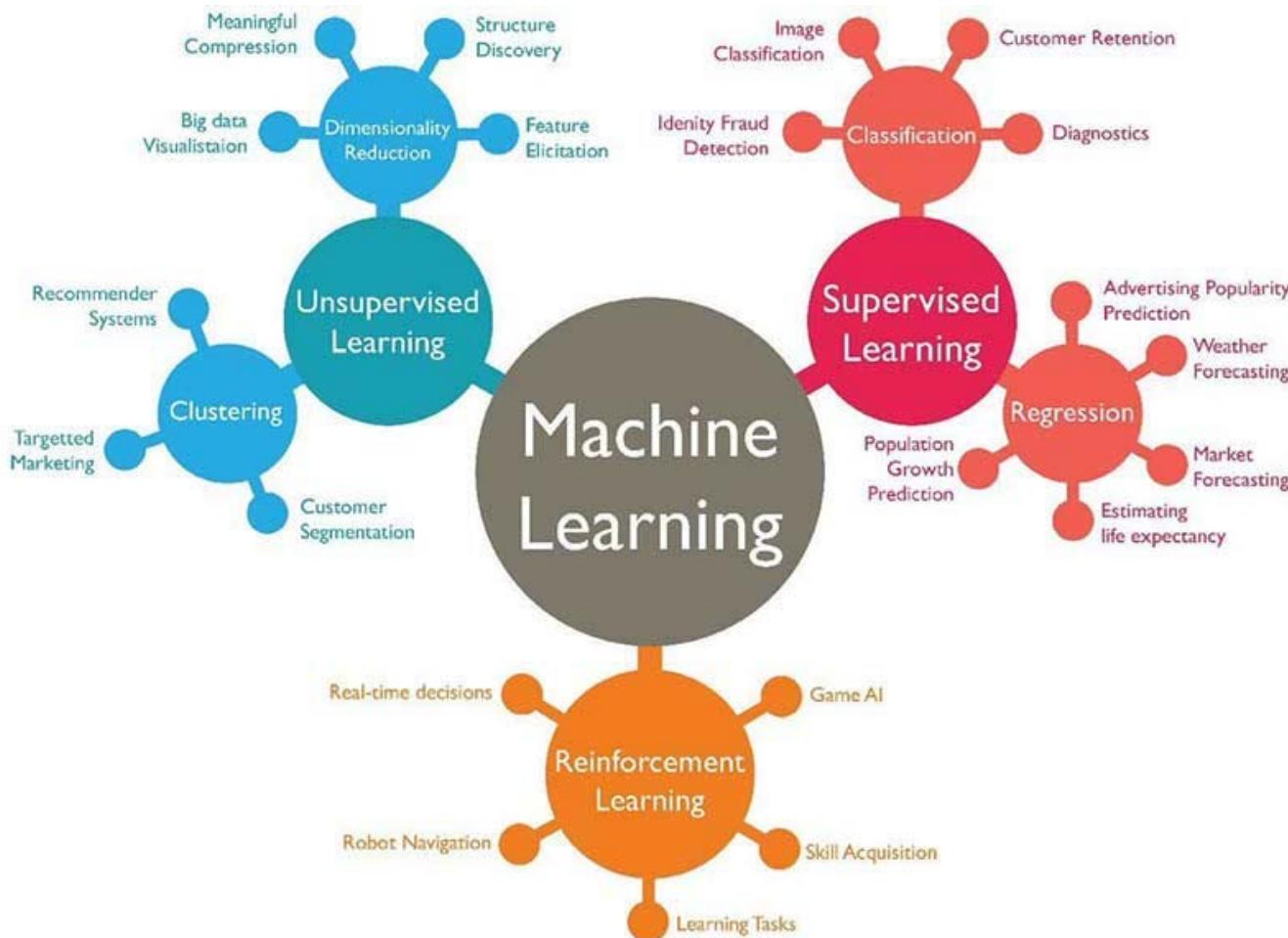


## فهرست

- مقدمه و کاربردهای یادگیری ماشین
- مثال یادگیری ماشین: دسته‌بندی دو نوع ماهی
- برخی مفاهیم یادگیری ماشین
- برخی چالش‌های یادگیری ماشین
- ارزیابی سیستم‌های یادگیری ماشین
  - ارزیابی سیستم‌های یادگیری باناظر
  - ارزیابی سیستم‌های یادگیری بی ناظر

# یادگیری ماشین . . .

## ○ ایجاد قابلیت یادگیری در کامپیووترها

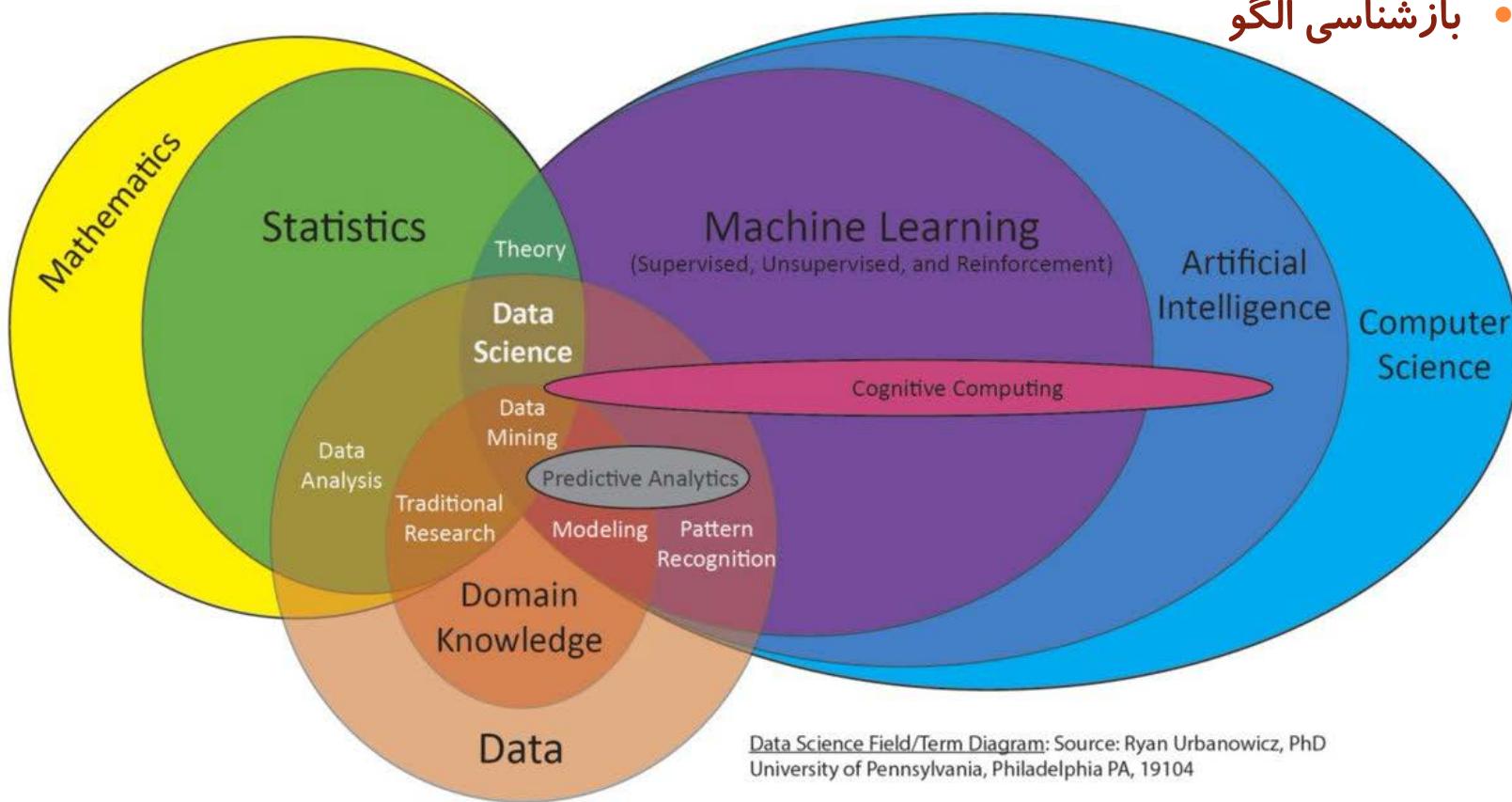




## یادگیری ماشین . . .

### ○ سایر مفاهیم مرتبط

- داده کاوی: کشف الگو در داده های موجود
- بازشناسی الگو





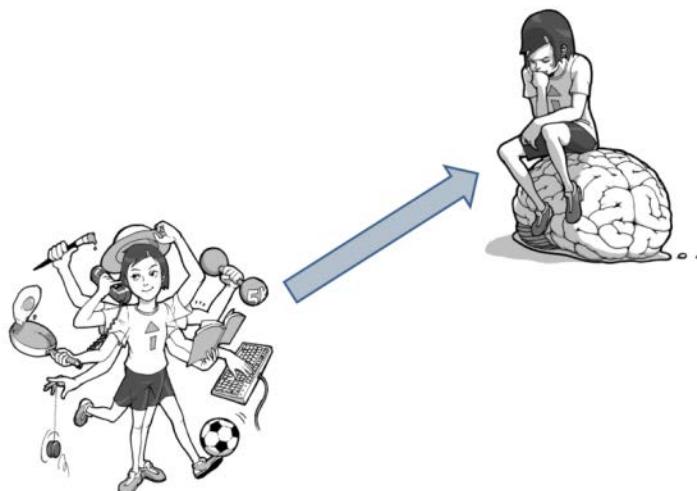
## یادگیری ماشین . . .

### ○ تشخیص (شناسایی، شناخت) (Cognition)

- نخستین مواجهه انسان با یک مسئله برای شناخت
- استفاده از تجربیات خود و دیگران
- ایجاد دسته (طبقه) جدید

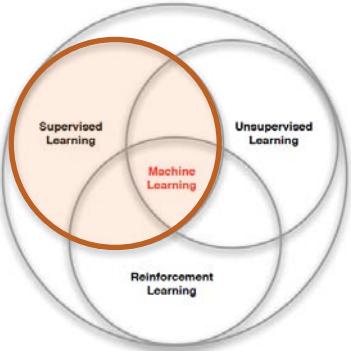
### ○ بازشناسی (Recognition)

- تشخیص (شناسایی) الگوهایی که قبلاً دیده شده‌اند
- الگوها کاملاً یا تقریباً مشابه الگوهای قبلی
- طبقه‌بندی الگوها در دسته‌های مجزا





# یادگیری با نظارت ...



## ○ یادگیری با نظارت (Supervised)

- داده‌های آموزش دارای برچسب هستند. برچسب‌ها بیانگر دسته هر داده هستند.
- مثال: دو دسته: سگ و گربه

### (Classification)

دسته: سگ



دسته: گربه





## یادگیری با نظرات ...

### ○ بازشناسی الگو (Pattern Recognition)

دریافت داده خام

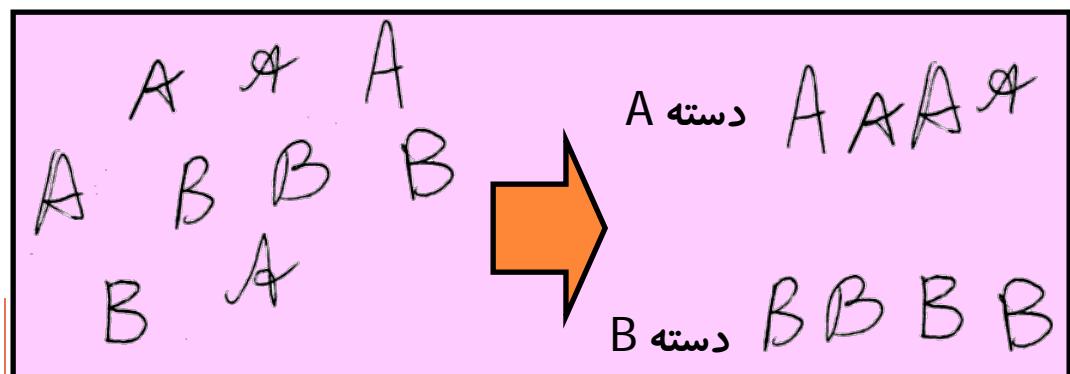
انجام یک عمل

بر اساس دسته‌بندی الگوهای آن داده

### ○ دسته‌بندی(طبقه‌بندی)/رگرسیون (Classification/Regression)

نوعی از بازشناسی برای دسته‌بندی الگو

دسته‌ها از قبل مشخص

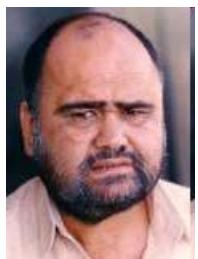




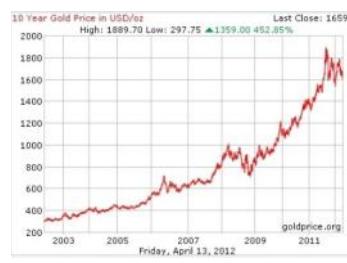
## یادگیری با نظارت ...

### ○ الگو (Pattern)

- در مقابل آشفتگی و هرج و مرج (chaos)
- یک شی مجرد (مانند مجموعه‌ای از اندازه‌ها که پدیده‌ای فیزیکی را تشریح می‌کند)
- معمولاً دارای اسم است



من در راه دیگر  
که نمایند این و شهاد رفته اند اما هیبت کنند  
که زمان ، نارا با خود بروند و شهاد نمایند .



### ○ مثال

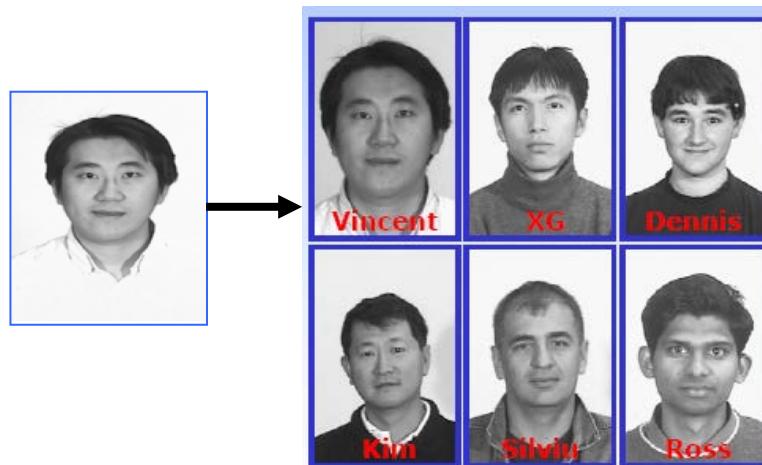
- تصویر (یک چهره)
- دستنوشته
- گفتار (صوت)
- متن
- اثر انگشت
- بارکد
- ....



## یادگیری با نظارت (کاربردها) ...

### ○ بازشناسی چهره (Face Recognition)

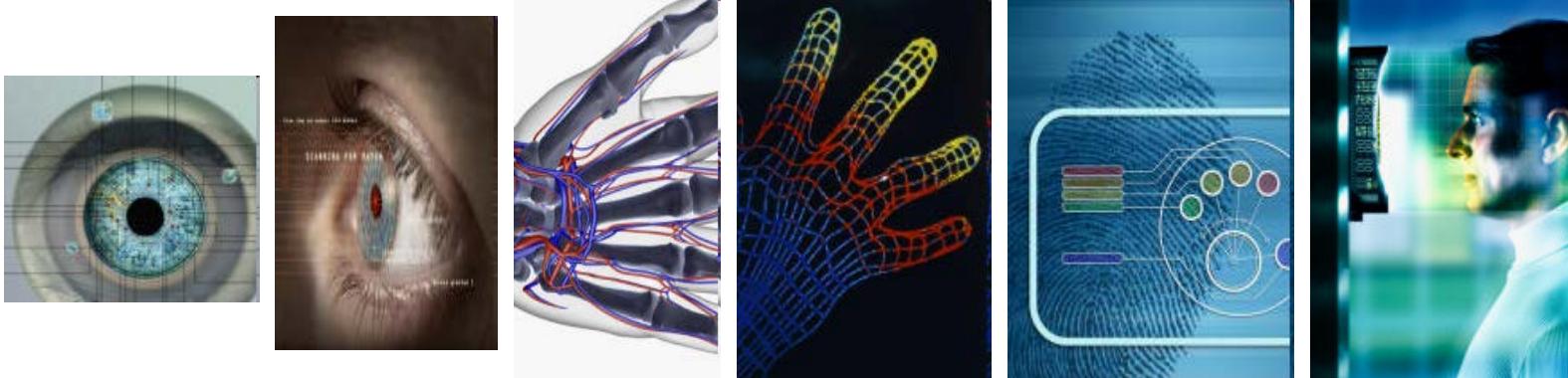
- کنترل دسترسی
- کاربردهای امنیتی مختلف





# یادگیری ماشین: یادگیری با نظارت (کاربردها) . . .

## روش‌های زیست‌سنجی (Biometric)





## یادگیری با نظارت (کاربردها) . . .

### • بازشناسی (Recognition)

- کشف (Detection): تشخیص محل چهره/پلاک (بدون شناسایی)

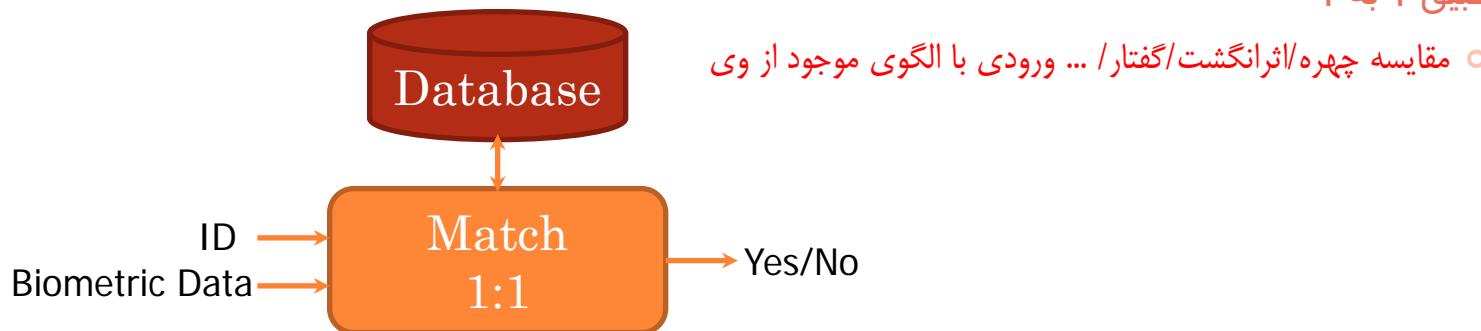


# یادگیری با نظارت (کاربردها) ...

## ○ بازشناسی (Recognition)

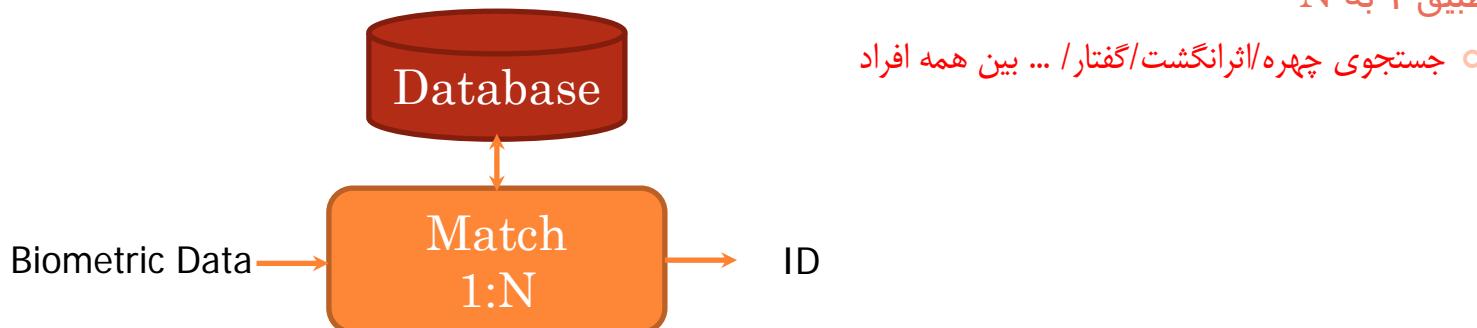
- تایید (Verification): تایید یا رد ادعای فرد مبنی بر اینکه کی هست

### ○ تطبیق ۱ به ۱



- شناسایی (Identification): تشخیص یک فرد از میان مجموعه‌ای از افراد

### ○ تطبیق ۱ به N



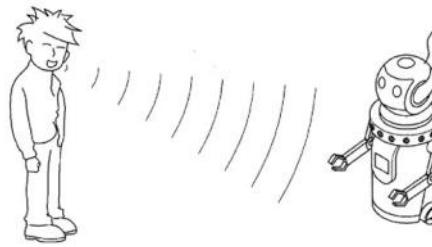
- جستجوی چهره/اثرانگشت/گفتار / ... بین همه افراد



## یادگیری با نظارت (کاربردها) ...

### ○ بازشناسی گفتار (Speech Recognition)

- تبدیل گفتار به متن: تایپ گفتاری، فرمان صوتی



### ○ بازشناسی گوینده (Speaker Recognition)

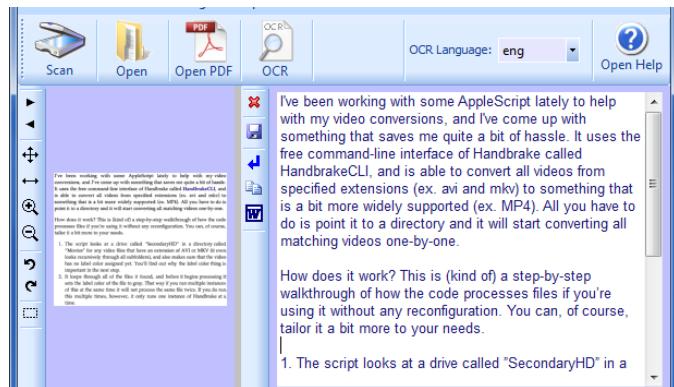
- تشخیص هویت از روی صدا (Identification)
- تایید هویت از روی صدا (Verification)



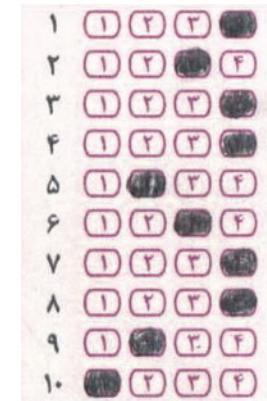
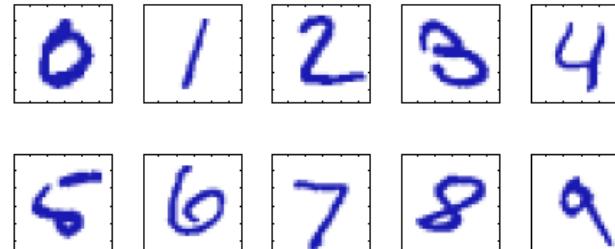
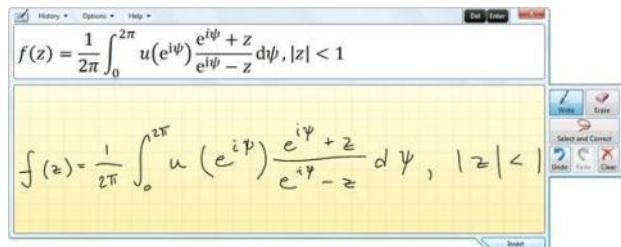
# یادگیری با نظارت (کاربردها)...

## ● بازشناسی تصویر (۱ از ۲)

- بازشناسی نویسه‌های نوری (OCR)
- بازشناسی دست خط (Handwritten Recognition)



عن نویسه‌های من  
موده هم زیبا می‌نمایم



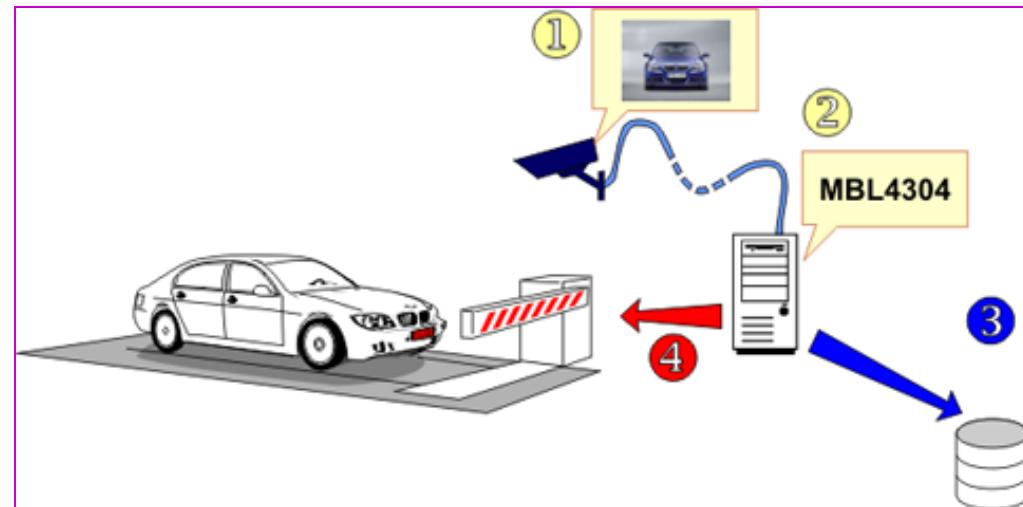
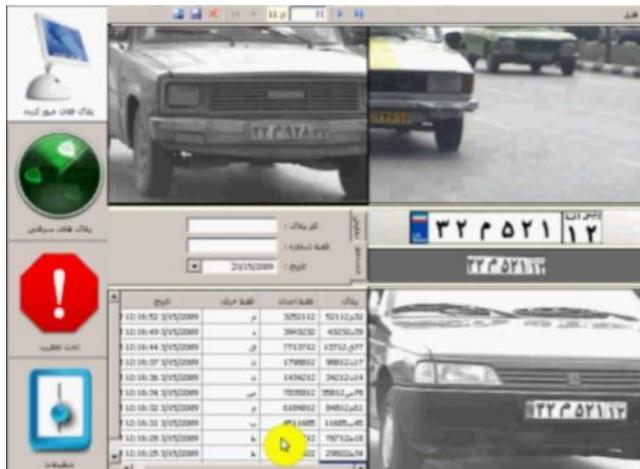


## یادگیری با نظارت (کاربردها) ...

### ● بازشناسی تصویر (۲ از ۲)

#### ● بازشناسی پلاک خودرو (License Plate Recognition: LPR)

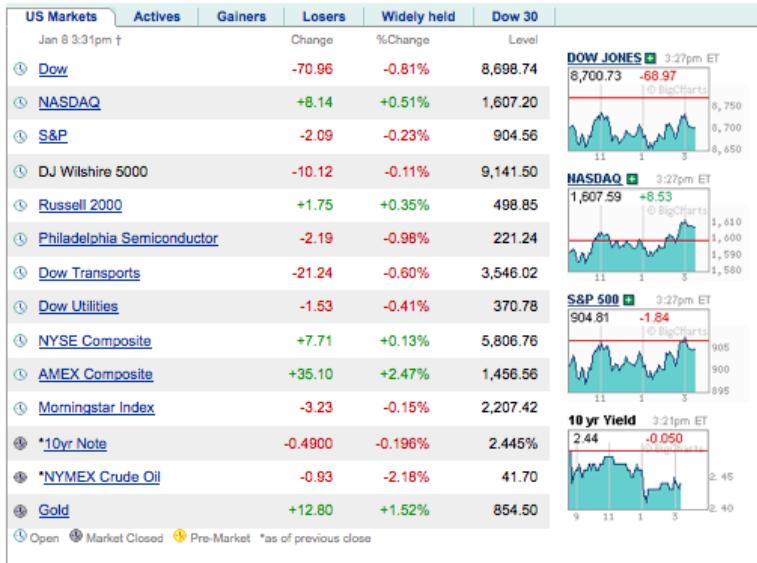
- کنترل‌های راهنمایی و رانندگی (سرعت غیرمجاز، ورورد به طرح ترافیک و ....)
- کنترل ورود/خروج پارکینگ‌ها
- پرداخت غیرنقدی عوارض





# یادگیری با نظارت (کاربردها) . . .

## ○ تحلیل‌ها و پیش‌بینی‌های مالی و تجاری



- پیش‌بینی بورس
- پیش‌بینی قیمت طلا

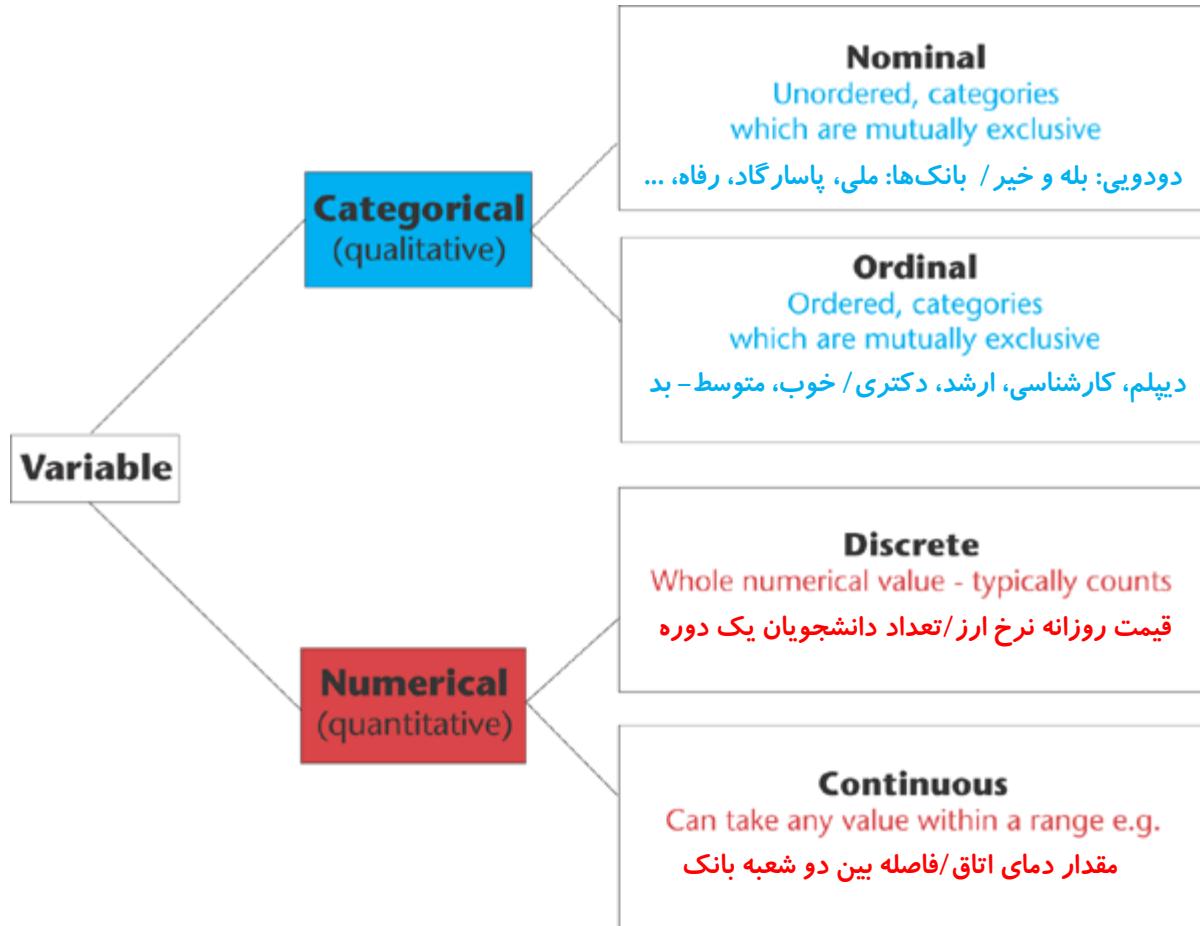
## ○ پژوهشی

- تشخیص بیماری‌ها از روی تصاویر و سیگنال‌ها



# یادگیری با نظارت . . .

## ○ انواع داده‌ها





# یادگیری با ناظارت ...

## ○ مساله

- ورودی: متغیرهای  $X = X_1, X_2, \dots, X_p$
- مثلاً یک عکس از چهره یک نفر / دمای هوا
- خروجی: متغیر  $y$
- مثلاً اسم صاحب عکس / میزان مصرف برق

- مساله = تخمین تابع  $y = f(X)$
- تشخیص / پیش‌بینی  $y$  از روی  $X$

## ○ تفاوت رگرسیون و دسته‌بندی

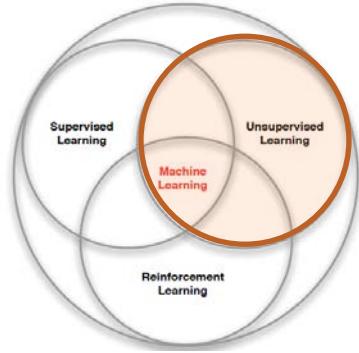
- اگر  $y$  یک متغیر عددی باشد  $\Leftrightarrow$  رگرسیون (regression)
- اگر  $y$  یک متغیر اسمی باشد  $\Leftrightarrow$  دسته‌بندی (classification)



## یادگیری با ناظارت

### ○ روش‌ها

- بیز ساده (NB: Naïve Bayes)
- نزدیک‌ترین همسایه (KNN: K Nearest Neighborhood)
- رگرسیون خطی / لجستیک (Linear/Logistic Regression)
- درخت تصمیم (DT: Decision Tree)
- ماشین بردار پشتیبان (SVM: Support Vector Machine)
- شبکه عصبی مصنوعی (ANN: Artificial Neural Network)
- مدل مخفی مارکوف (HMM: Hidden Markov Model)
- ... •



# یادگیری بدون نظارت ...

## ○ یادگیری بدون نظارت (Unsupervised)

- داده‌های آموزش دارای برچسب نیستند.
- خوشه‌بندی (Clustering) •

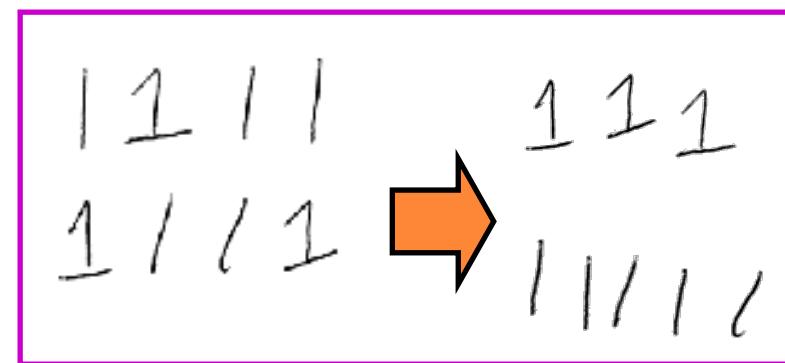


# یادگیری بدون ناظارت ...

## ○ خوشبندی (Clustering)

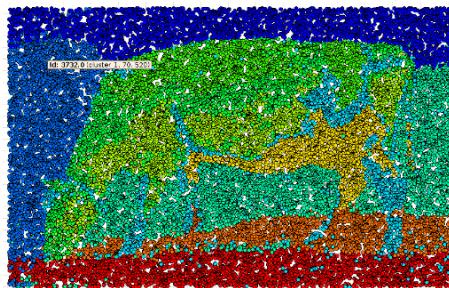
- دسته‌بندی الگوهای بر اساس شباهت

- دسته‌ها از قبل نامشخص



## پادگیری بدون نظارت ...

I have a Dream - Martin Luther King Jr.



## کاربردہا

- دسته‌بندی خودکار داده‌های جیم
  - متون وب‌سایت
  - رفتار مشتری‌ها برای بازیابی هدفمند
  - داده‌کاوی و استخراج دانش از داده‌های
  - کاهش ابعاد ویژگی‌ها
  - قطعه‌بندی تصویر (Segmentation)



# یادگیری بدون نظارت

## روش‌ها

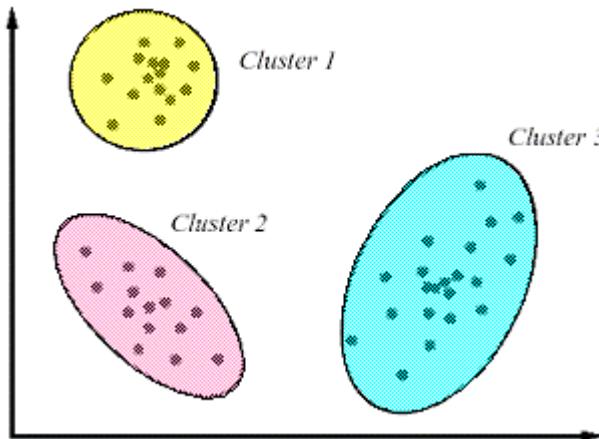
- شبکه عصبی

- نگاشت‌های خودسازمانده کوهون (SOM)

- نظریهٔ نوسان وفقی (ART)

- K میانگین (K-means) و حالت فازی آن (FCM)

- روش‌های سلسله مراتبی





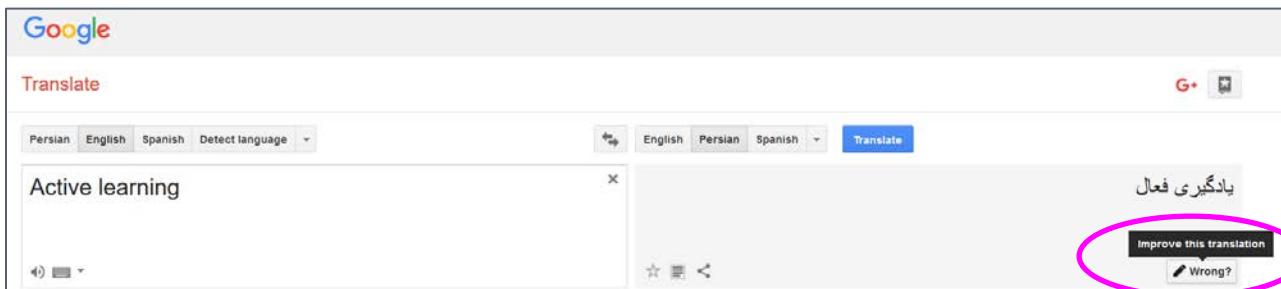
## یادگیری نیمه نظارتی

### ○ یادگیری نیمه نظارتی (Semi-Supervised)

- بخش (کمی) از داده برچسب دارد و بخش (عمده‌ای) برچسب ندارد
- روش Boot-Strapping
- آموزش دسته‌بند با داده‌های دارای برچسب (در ابتدا با حجم کم)
- برچسب زدن خودکار به سایر داده‌ها (ی بدون برچسب) با دسته‌بند موجود
- تکرار دو گام قبل (آموزش مجدد دسته‌بند و برچسب زنی مجدد)

### ○ یادگیری فعال (Active Learning)

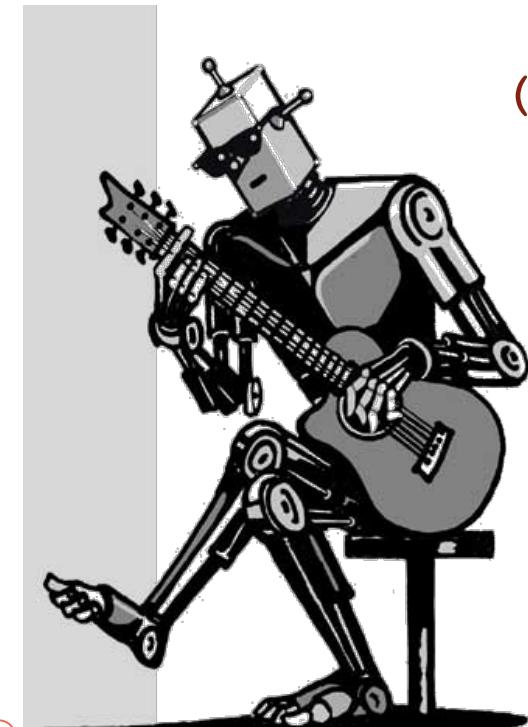
- آموزش دسته‌بند با داده‌های دارای برچسب (در ابتدا با حجم کم)
- درخواست از کاربران برای برچسب زدن به نمونه داده‌ها برای افزایش حجم داده





## یادگیری تقویتی ...

### ... (Reinforcement Learning) یادگیری تقویتی

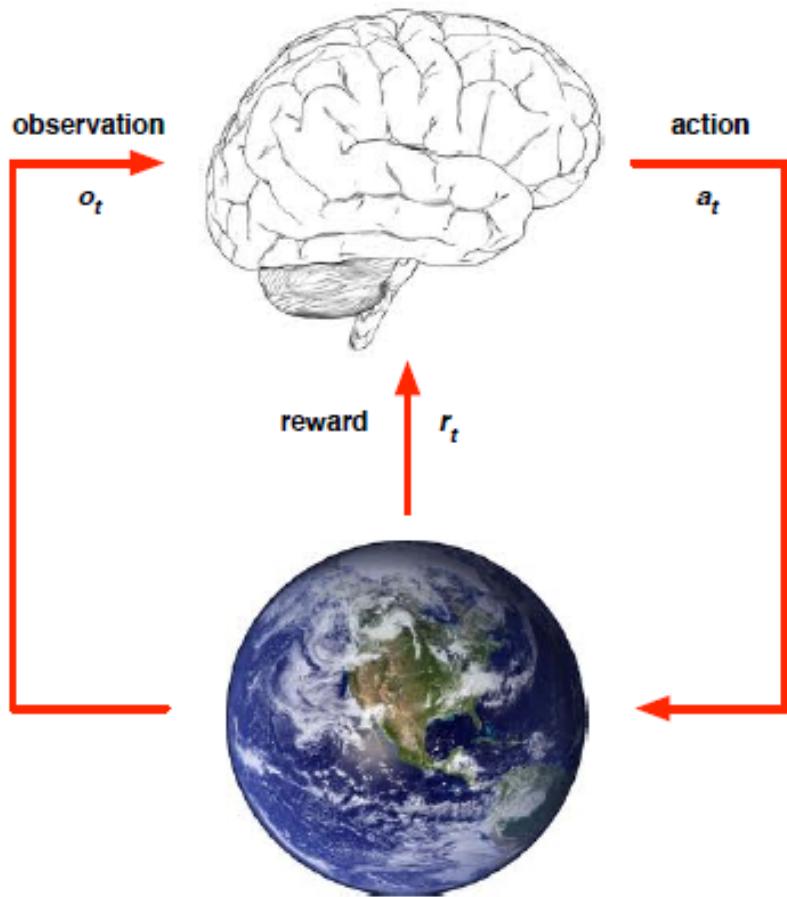


- یادگیری بر مبنای آزمون و خطا در تعامل با محیط

- انجام یک عمل (action) و دریافت بازخورد از محیط (Penalty)
- یادگیری بر اساس پاداش (Reward) و جریمه (supervisor)
- عدم وجود ناظر (supervisor)
- یادگیری در طول زمان (عدم استقلال مشاهدها)

# یادگیری ماشین: یادگیری تقویتی . . .

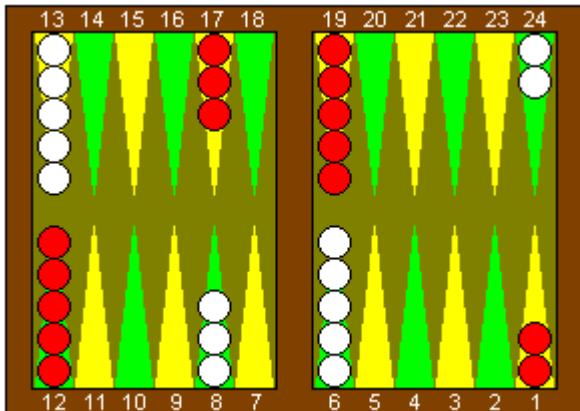
## ○ مراحل یک الگوریتم یادگیری تقویتی



- At each step  $t$  the agent:
  - Executes action  $a_t$
  - Receives observation  $o_t$
  - Receives scalar reward  $r_t$
- The environment:
  - Receives action  $a_t$
  - Emits observation  $o_t$
  - Emits scalar reward  $r_t$

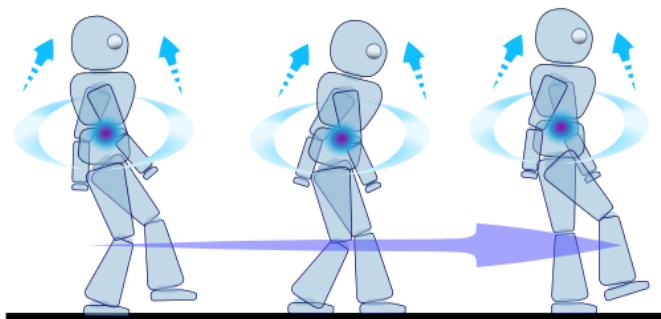


# یادگیری ماشین: یادگیری تقویتی . . .



## مثال

- یاد گرفتن بازی تخته نرد
  - پاداش = بردن بازی
  - جریمه = باختن بازی
  - مثال
    - با پاداش +100 برای برد و -100 برای باخت
    - بعد از ۱.۵ میلیون بازی در حد یک انسان خبره یاد می‌گیرد



- یاد گرفتن راه رفتن ربات
  - پاداش = حرکت رو به جلو
  - جریمه = افتادن



# یادگیری ماشین: یادگیری تقویتی . . .

## ○ زمینه‌های مرتبط

- نظریه تصمیم (decision theory) در آمار و مدیریت
- نظریه کنترل (control theory) در مهندسی

## • نظریه بازی‌ها (game theory)

- حالت توسعه یافته یادگیری تقویتی
- عوامل در تعامل با محیط یاد می‌گیرند و پاسخ تولید می‌کنند اما  
◦ در یادگیری تقویتی محیط ثابت است
- در نظریه بازی سایر عوامل در محیط نیز متناسب اعمال ما پاسخ خود را تغییر می‌دهند



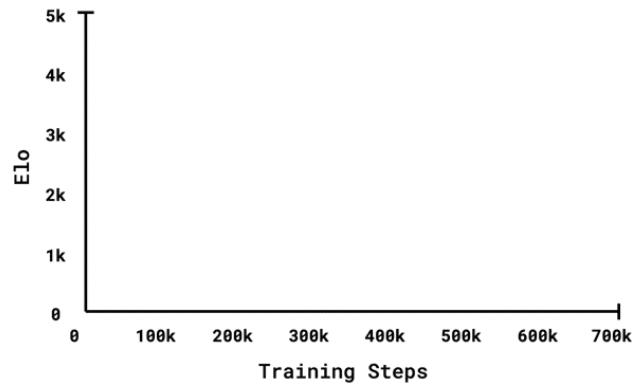
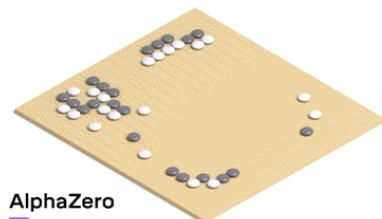
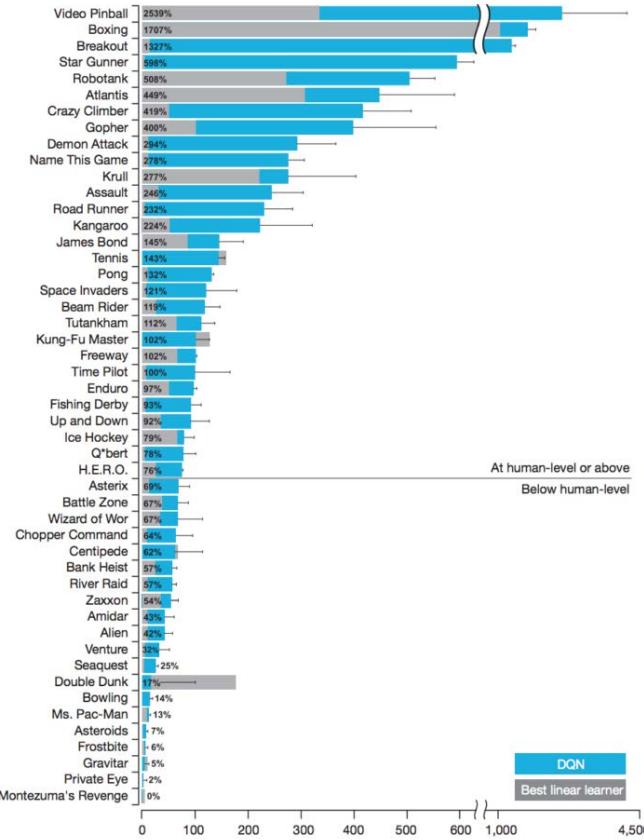


# یادگیری ماشین: یادگیری تقویتی

## پیروزی در بازی Go، Shogi و شترنج (مدل AlphaZero گوگل)

Reinforcement Learning •

CNN •



[Silver, David, et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play.", 2018]



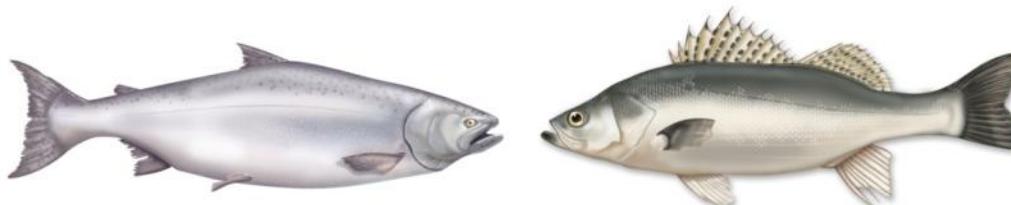
## مثال: دسته‌بندی دو نوع ماهی ...

### ○ مساله

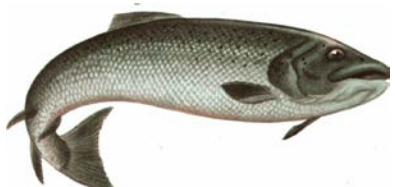
- ماهی‌های صید شده توسط یک قایق ماهی‌گیری دو نوع هستند:

◦ خارماهی یا باس (Sea Bass)

◦ ماهی سالمون (Salmon)



- ماهی‌ها برای فرآیندهای بعدی (نگهداری، قیمت‌گذاری و ....) باید به دو دسته تفکیک شوند





## مثال: دسته‌بندی دو نوع ماهی ...



### راه حل ...

- استفاده از یک تسمه نقاله متحرک
- نصب یک دوربین بالای تسمه نقاله
- گذاشتن ماهی‌های صید شده بر روی تسمه نقاله

- گرفتن یک عکس از هر ماهی
- پردازش عکس گرفته شده و حذف بخش‌های اضافی از تصویر ( فقط تصویر ماهی بماند)
- استخراج برخی اطلاعات از تصویر ماهی ( اندازه، رنگ، تعداد باله، شکل سر و دم و ... )
- پردازش اطلاعات اخذ شده و تصمیم‌گیری در مورد آن

دریافت داده

Data Acquisition

پیش‌پردازش

Preprocessing

استخراج ویژگی

Feature Extraction

دسته‌بندی

Classification

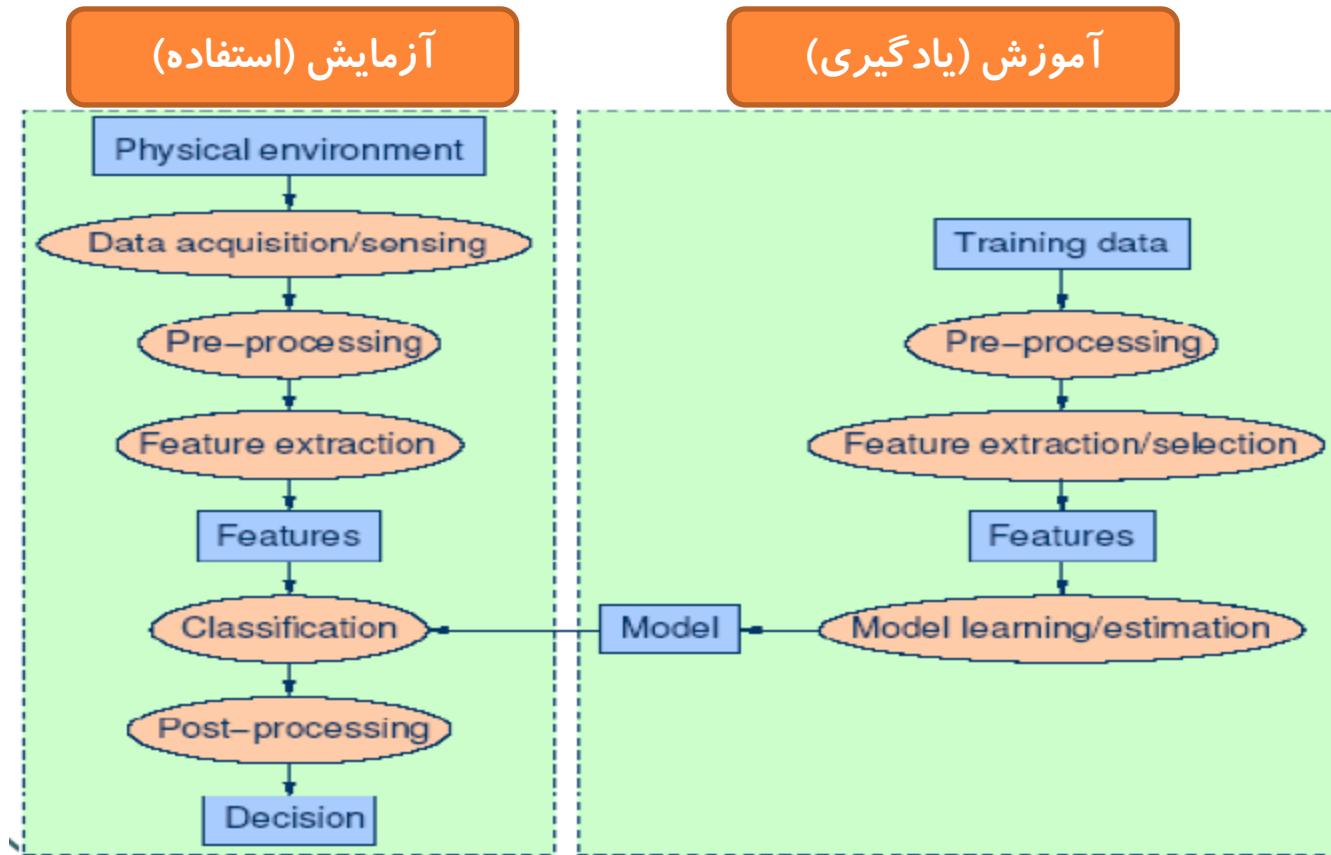


## مثال: دسته‌بندی دو نوع ماهی ...

- سوال: دسته‌بندی چگونه صورت می‌گیرد؟
- یک سیستم بازشناسی الگو شامل دو مرحله (بخش) است:
  - آموزش (یادگیری)
    - یادگیری الگوهای (ویژگی‌های الگوهای) موردنظر
    - از روی نمونه داده‌ها (در این مثال، از روی تصاویر مختلف از دو نوع ماهی)- **مجموعه آموزش**
    - ایجاد مدل برای هر کدام از دسته‌ها (در این مثال، دو نوع ماهی)
  - آزمایش (استفاده)
    - استفاده از مدل ساخته شده در مرحله آموزش
    - اندازه‌گیری شباهت نمونه ناشناس ورودی (در این مثال، ماهی روی تسمه نقاله) به مدل‌های موجود

# مثال: دسته‌بندی دو نوع ماهی ...

نمای کلی یک سیستم بازشناسی الگو

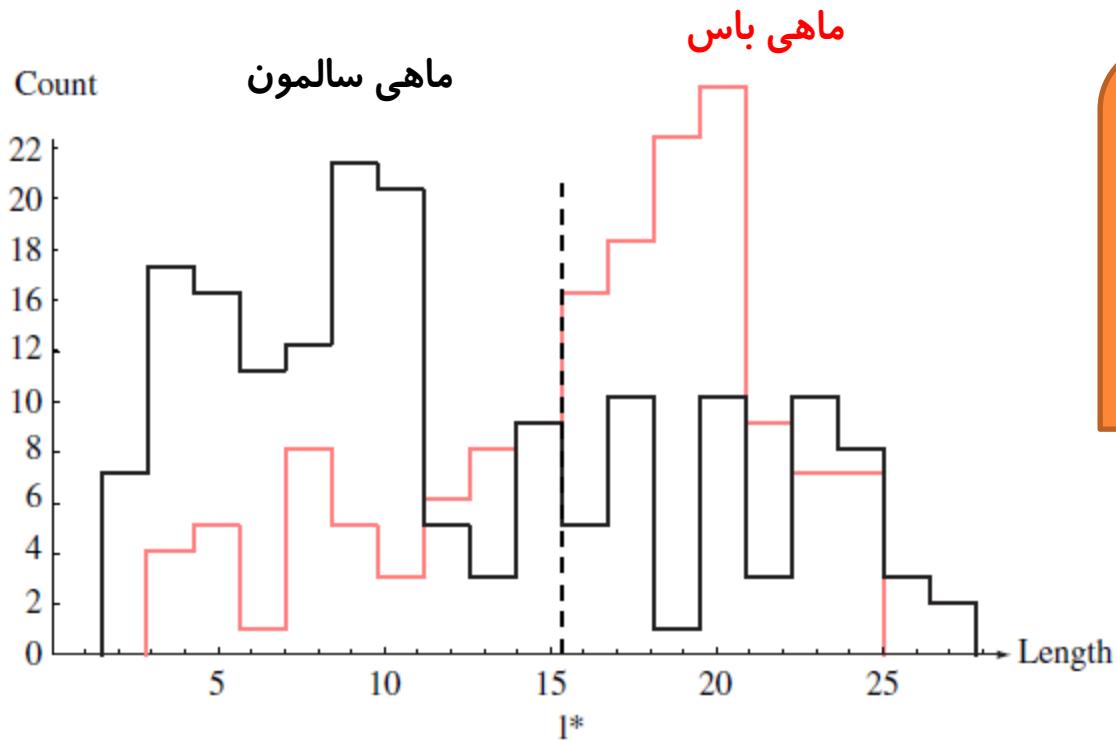




## مثال: دسته‌بندی دو نوع ماهی ...

### استخراج ویژگی: طول ماهی

- محاسبه طول تعدادی از ماهی‌های باس و تعدادی از ماهی‌های سالمون
- هیستوگرام فرآوانی طول دو نوع ماهی



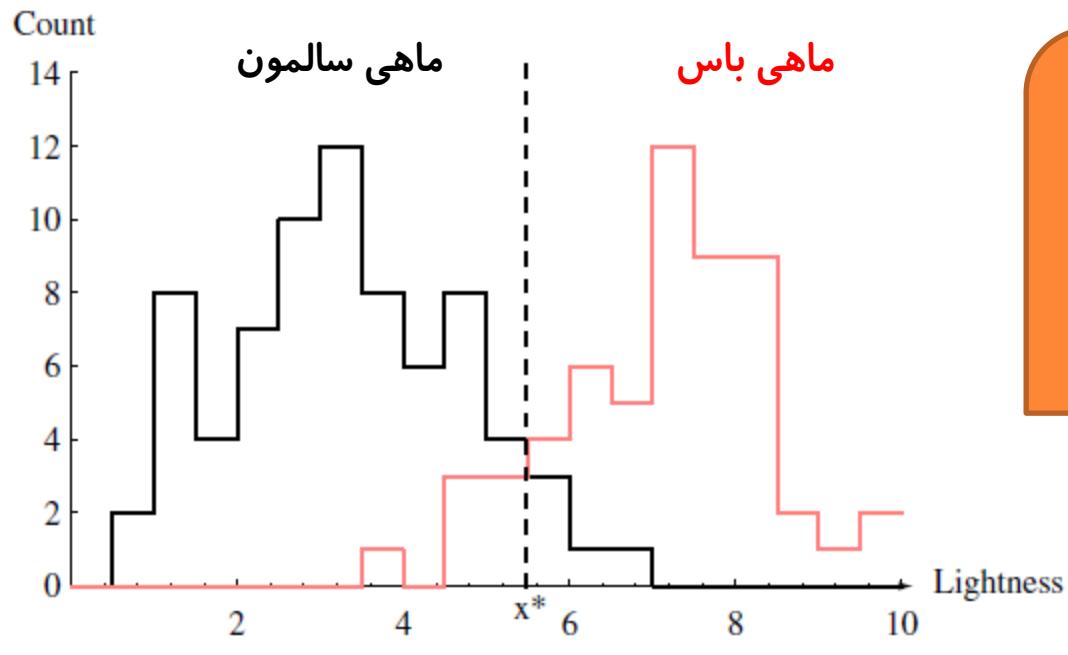
هرچند به طور متوسط طول ماهی باس بیشتر از سالمون است اما همیشه این گونه نیست!



## مثال: دسته‌بندی دو نوع ماهی ...

### استخراج ویژگی: وزن ماهی

- محاسبه وزن تعدادی از ماهی‌های باس و تعدادی از ماهی‌های سالمون
- هیستوگرام فرآوانی وزن دو نوع ماهی



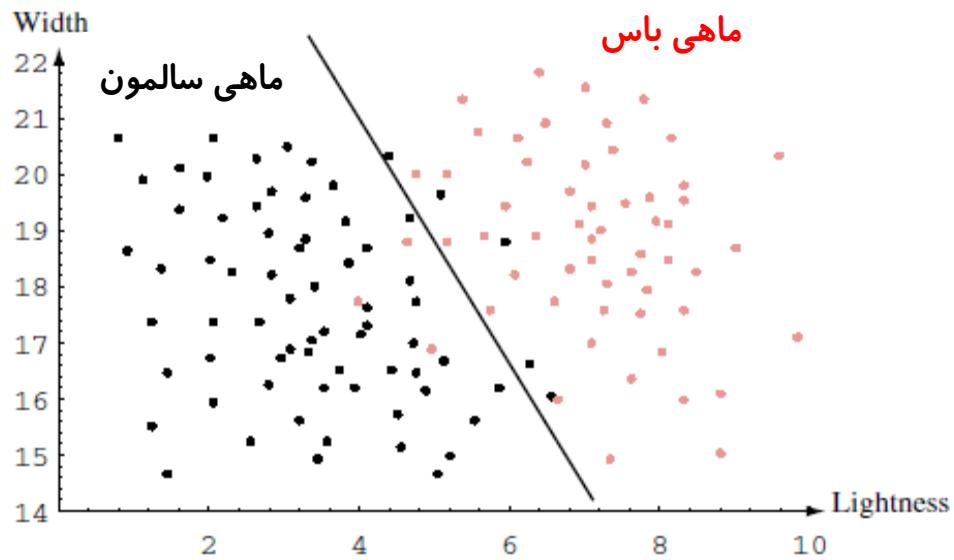
میزان همپوشانی مقادیر وزن دو نوع ماهی از همپوشانی طول آنها کمتر است = ویژگی بهتر



## مثال: دسته‌بندی دو نوع ماهی ...

### استخراج ویژگی: طول + وزن

- در کاربردهای واقعی از بیش از یک ویژگی استفاده می‌شود
- $x_1 = \text{طول}$  و  $x_2 = \text{وزن}$
- استفاده همزمان از دو ویژگی فوق = فضای دو بعدی (2D)
- بردار ویژگی =  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$



تفکیک بیشتر فضا با  
استفاده از دو ویژگی  
همزمان



## مثال: دسته‌بندی دو نوع ماهی ...

### ○ استخراج ویژگی: بیش از یک ویژگی

- می‌توان از ویژگی‌های سوم (نسبت طول به عرض)، چهارم (تعداد باله) و ... استفاده کرد.

### • سوال

- چه تعداد ویژگی لازم است؟
- چگونه به مفید (خوب) بودن یک ویژگی پی ببریم؟
- وقتی تعداد ویژگی‌ها زیاد باشد، تعدادی از آنها با هم همبستگی داشته و حاوی اطلاعات مشابه هستند.  
در این حالت چه اتفاقی می‌افتد؟

### ○ استخراج ویژگی به نوع مساله و راه حل انتخابی بستگی دارد

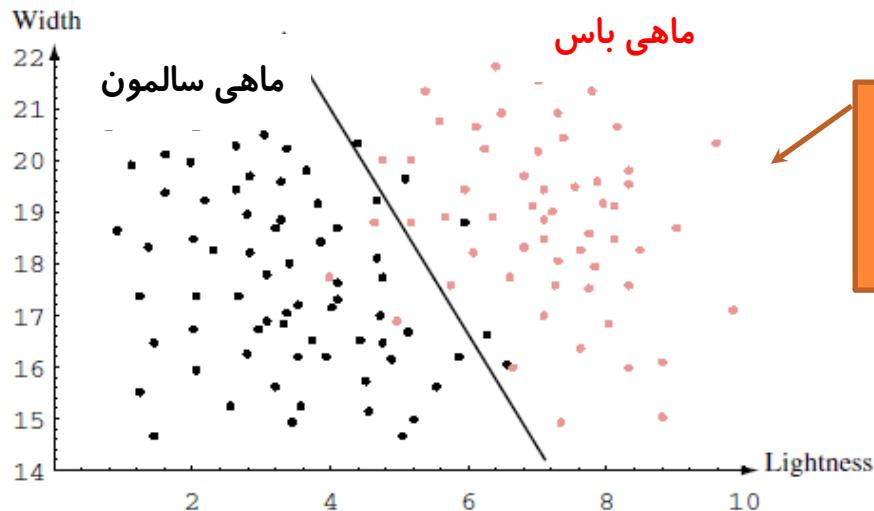
### ○ استخراج ویژگی مناسب

- استخراج آسان
- قدرت تفکیک بالا و افزونگی اندک
- تعداد پایین و پردازش آسان



## مثال: دسته‌بندی دو نوع ماهی ...

### ... (Decision Boundary)

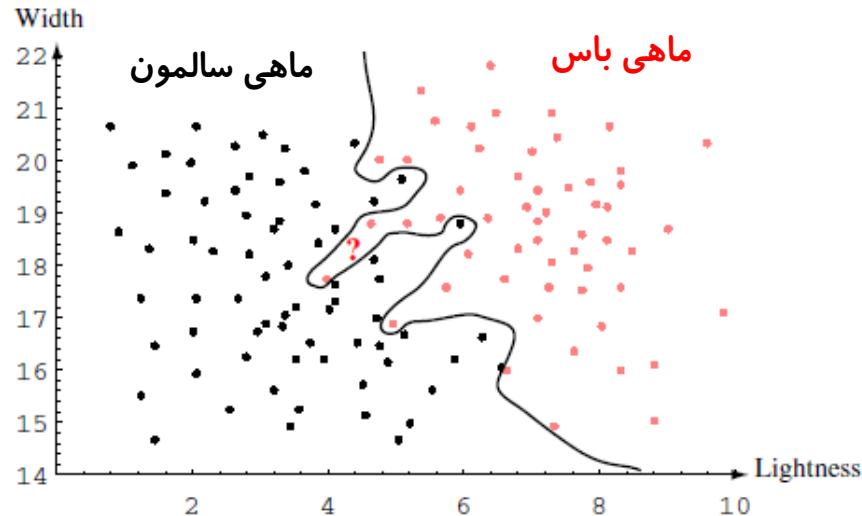


**مرز خیلی پیچیده**

- دسته‌بندی درست تمام (غلب) نمونه‌های آموزش
- تعیین‌پذیری (Generalization) پایین

- مرز تفکیک کننده دسته‌های مختلف

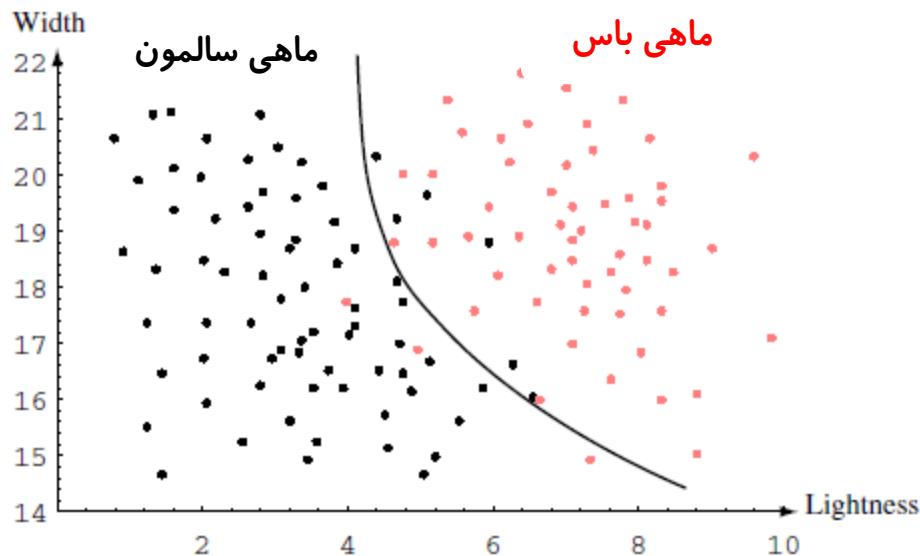
مرز ساده = خطی  
عدم دسته‌بندی درست برخی از نمونه‌ها



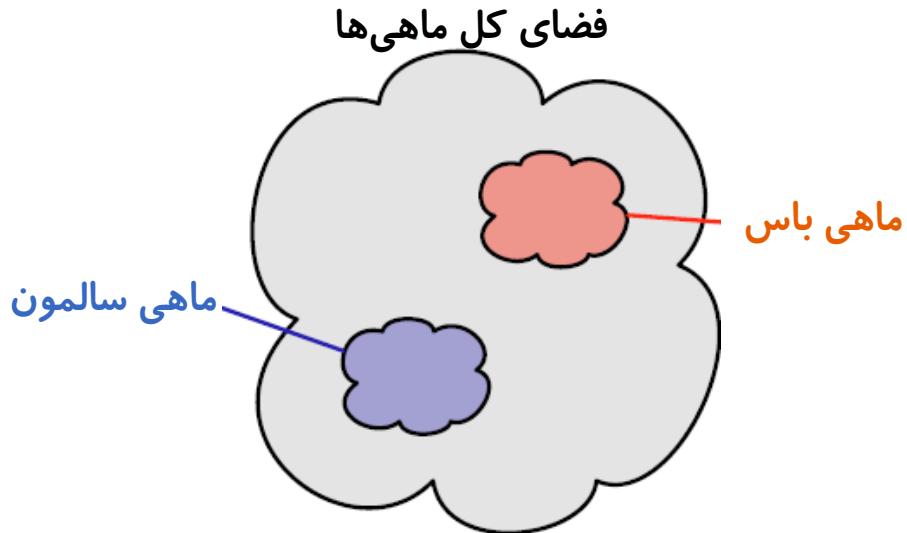
## مثال: دسته‌بندی دو نوع ماهی ...

### مرز تصمیم‌گیری (Decision Boundary)

- راه حل میانی = مرز تصمیم‌گیری نه چندان پیچیده
- قابلیت تعمیم‌پذیری (Generalization) بالا
- پاسخ‌گویی درست شبکه به داده‌های جدید

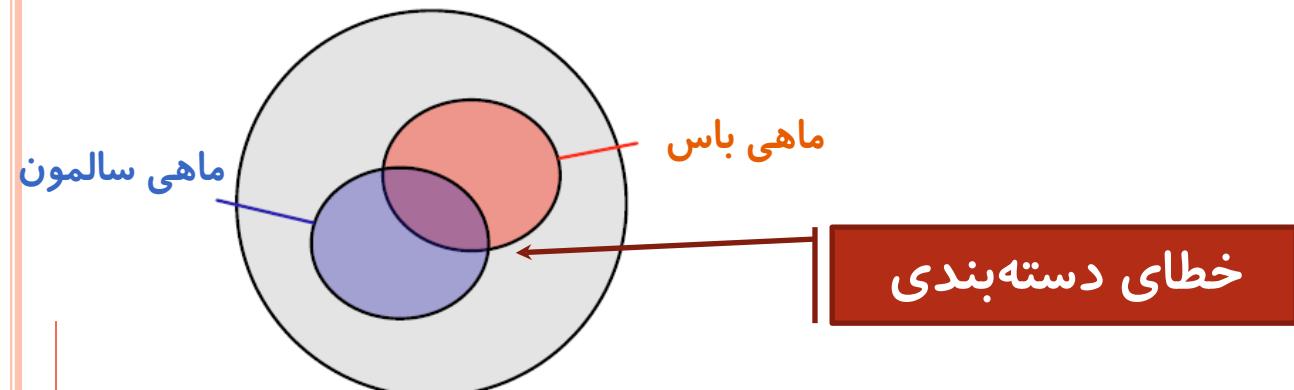


## مثال: دسته‌بندی دو نوع ماهی



- فضای حالت ایده‌آل

- فضای حالت در عمل (ایجاد شده با ویژگی‌های استخراج شده)





# رویکردهای مختلف در یادگیری ماشین

## ○ آماری (Statistical)

- مدل کردن اطلاعات (ویژگی‌ها) الگوها با روش‌های آماری
- روش: بیز ساده، HMM

## ○ تطبیق الگو (Template Matching)

- مقایسه الگوها بر اساس یک معیار شباهت (نوعی همبستگی)
- روش: DTW

## ○ مبتنی بر قاعده (Syntactic)

- طراحی دسته‌بندها بر اساس مجموعه‌ای از قواعد
- روش: درخت تصمیم

## ○ مبتنی بر محاسبات نرم

- شبکه‌های عصبی مصنوعی (Artificial Neural Networks)

## ○ ترکیبی (Hybrid)

- Mixture of Experts و Super Classifier



## بهترین دسته‌بندی کننده چیست؟

### No Free Lunch Theorem ◉

- Wolpert and Macready (2005) "any two optimization algorithms are equivalent when their performance is averaged across all possible problems"

• به صورت کلی هیچ دسته‌بندی کننده نیست که همیشه بهترین باشد.

• از ناهار مجانی خبری نیست!

◦ تعدادی رستوارن با منوی غذای ثابت و قیمت‌های متفاوت

◦ توجه به عادت غذایی (گوشت‌خوار یا گیاه‌خوار) = یافتن رستوران با غذای ارزان

◦ فرد همه‌چیز‌خوار = دفعات زیادی به رستوران می‌رود = تفاوتی ندارد به کدام رستوران برود = به طور متوسط هزینه غذا تقریباً ثابت است

◦ روش دسته‌بندی = رستوران؛ مسائل = غذا؛ هزینه دسته‌بندی = هزینه غذا

◦ هزینه حل تمام مسائل با هر یک از روش‌ها، ثابت است.

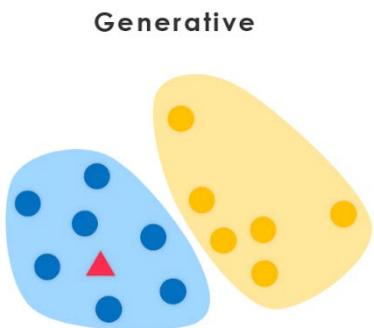




## مفاهیم ...

### ○ یادگیری/آموزش تولیدی (Generative)

- یادگیری مدل یک دسته بر اساس آنچه نمونه داده‌های آن از آن تولید شده است



- یادگیری توزیع احتمال داده‌ها

- فقط به داده‌های یک دسته توجه می‌کند
- استفاده از تابع احتمال توام داده و دسته

روش‌ها: HMM، DBN، RBM و GAN

### ○ یادگیری/آموزش تمایزی (Discriminative)

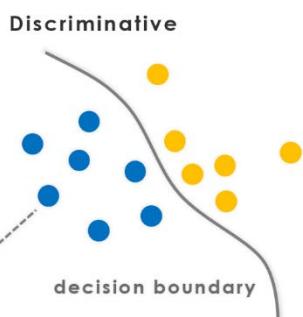
- یادگیری تمایز بین دسته‌ها ← توجه به داده‌های همه دسته‌ها در یادگیری

- یادگیری مرز بین دسته‌ها

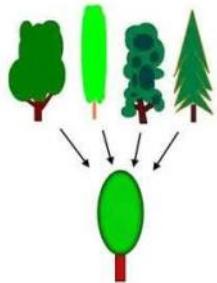
- استفاده از تابع احتمال شرطی دسته به شرط داده

- عدم یادگیری توزیع داده‌ها

روش‌ها: SVM، MLP

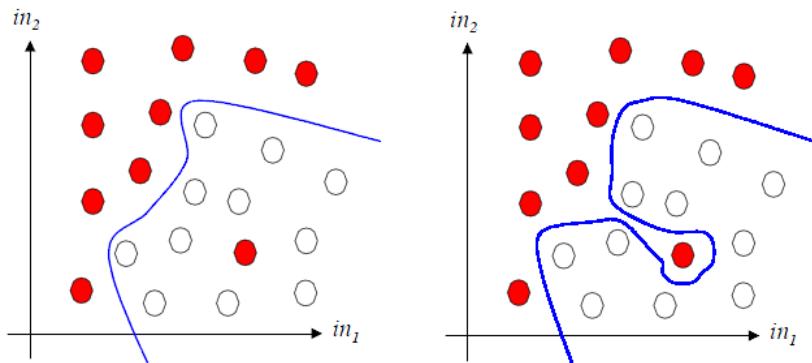


## مفاهیم ...



### ○ تعمیم‌پذیری (Generalization)

- بازشناسی الگوهای ورودی جدید که با الگوهای آموخت داده شده (قبل‌اً مشاهده شده) تا حدودی متفاوت است



### ○ بیش برآزش (Overfitting)

- دسته‌بندی خیلی دقیق داده آموخت
- کاهش قابلیت تعمیم

### ○ نویز (Noise)

- هر گونه مشخصه ناخواسته‌ای که باعث عدم نمایش درست الگو یا ویژگی‌های آن باشد





## مفاهیم ...

### ○ خطای دسته‌بندی

$$\begin{aligned} E[(y - \hat{f})^2] &= E[y^2 + \hat{f}^2 - 2y\hat{f}] \\ &= E[y^2] + E[\hat{f}^2] - E[2y\hat{f}] \\ &= \text{Var}[y] + E[y]^2 + \text{Var}[\hat{f}] + E[\hat{f}]^2 - 2E[\hat{f}]E[y] \\ &= \text{Var}[y] + \text{Var}[\hat{f}] + (E[\hat{f}]^2 - 2E[\hat{f}]E[y] + E[y]^2) \\ &= \text{Var}[y] + \text{Var}[\hat{f}] + (E[\hat{f}] - E[y])^2 \\ &= \text{Var}[y] + \text{Var}[\hat{f}] + \text{Bias}[\hat{f}]^2 \\ &= \sigma^2 + \text{Var}[\hat{f}] + \text{Bias}[\hat{f}]^2 \end{aligned}$$

$$\text{Error}(x) = E \left[ (f(x) - \hat{f}(x))^2 \right] = (E[\hat{f}(x)] - f(x))^2 + E \left[ (\hat{f}(x) - E[\hat{f}(x)])^2 \right] + \sigma^2$$

متوجه خطا بین مقدار واقعی و  
مقدار تخمین زده شده

$$\text{Bias}(\hat{f}(x)) = E[\hat{f}(x) - f(x)]$$

پراکندگی (واریانس) مقادیر تخمینی  
برای یک نقطه (در صورت  
نمونه‌برداری‌های مختلف)

$$\text{Var}(\hat{f}(x)) = E[(\hat{f}(x) - E[\hat{f}(x)])^2]$$

$$\sigma^2 = \text{irreducible error} = \text{Var}(f(x))$$

$$\text{Error}(x) = \left( \text{Bias}(\hat{f}(x)) \right)^2 + \text{Var}(\hat{f}(x)) + \sigma^2$$

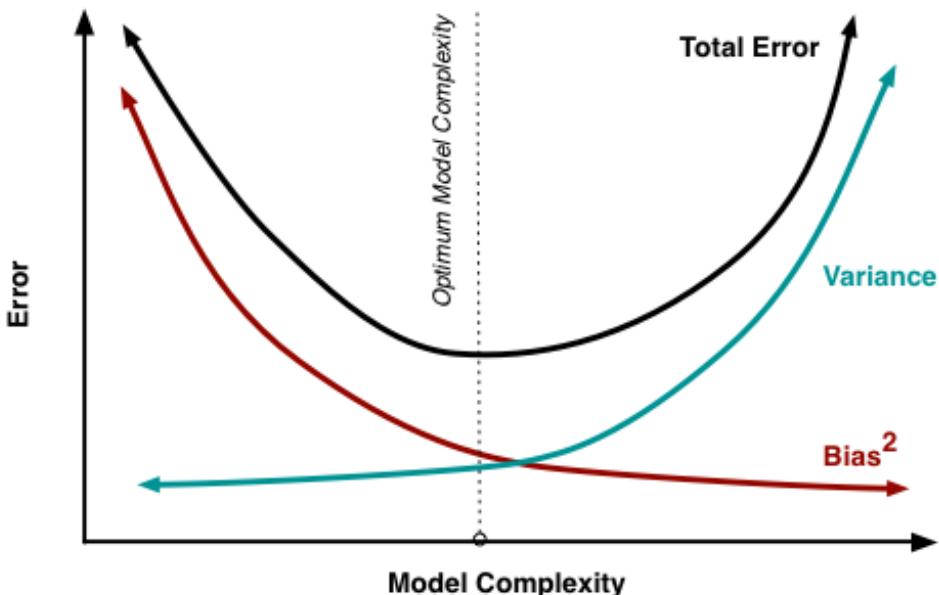
پراکندگی (واریانس) داده‌های اصلی

## مفاهیم ...

$$\text{Error}(x) = \left( \text{Bias}(\hat{f}(x)) \right)^2 + \text{Var}(\hat{f}(x)) + \sigma^2$$

- خطا: تعلق بین بایاس و واریانس
- کاهش یکی باعث افزایش دیگری می شود

- پیچیده کردن مدل = بیش برآورده = بایاس کمتر = واریانس بیشتر
- مدل ساده تر = بایاس بیشتر = واریانس کمتر

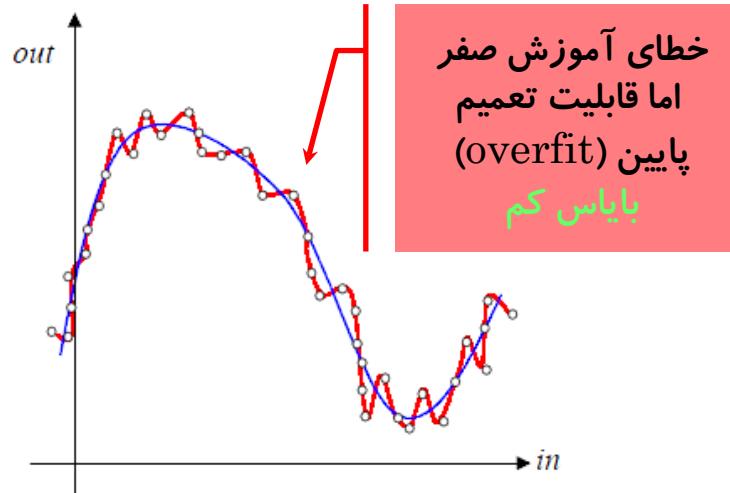
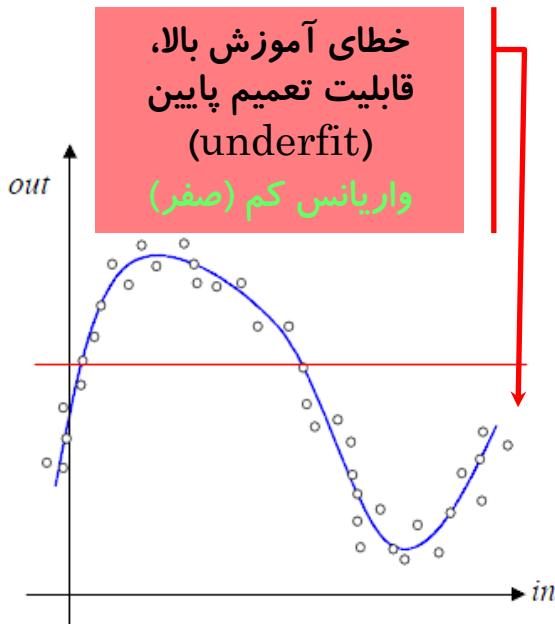


$$\frac{dBias}{dComplexity} = -\frac{dVariance}{dComplexity}$$

## مفاهیم ...

$$\text{Error}(x) = \left( \text{Bias}(\hat{f}(x)) \right)^2 + \text{Var}(\hat{f}(x)) + \sigma^2$$

### ○ خطای تعامل بین بایاس و واریانس



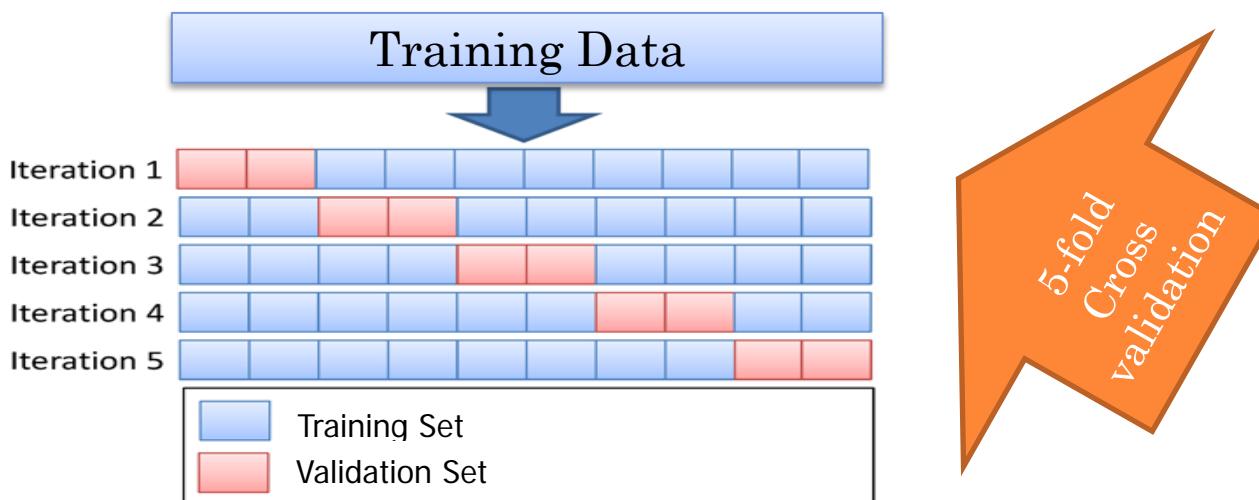
### ○ کم کردن واریانس داده‌ها = افزایش حجم داده‌ها

## مفاهیم

- تفکیک داده‌ها به دو مجموعه آموزش و آزمون
- آموزش ۸۰٪ و آزمون ۲۰٪ کل داده‌های موجود

### ○ تعادل بین یادگیری الگوها و تعمیم

- استفاده از دو مجموعه داده مجزا در زمان آموزش شبکه
- یک مجموعه برای آموزش الگوها و یک مجموعه برای آزمون الگوها (مجموعه validation)
- روش Cross Validation: تقسیم داده آموزش به  $K$  زیرمجموعه
- هر بار یکی از زیرمجموعه‌ها برای تایید اعتبار استفاده می‌شود

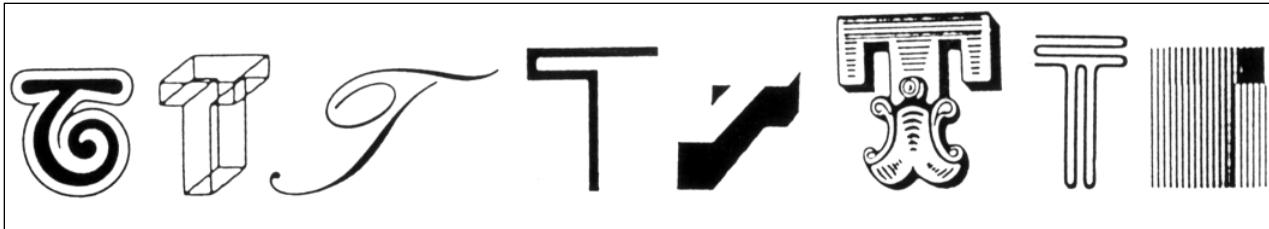




## چالش‌ها ...

### ○ تنوع درون دسته‌ای (Intra-Class Variability)

- دسته: حرف T



- دسته: یک چهره

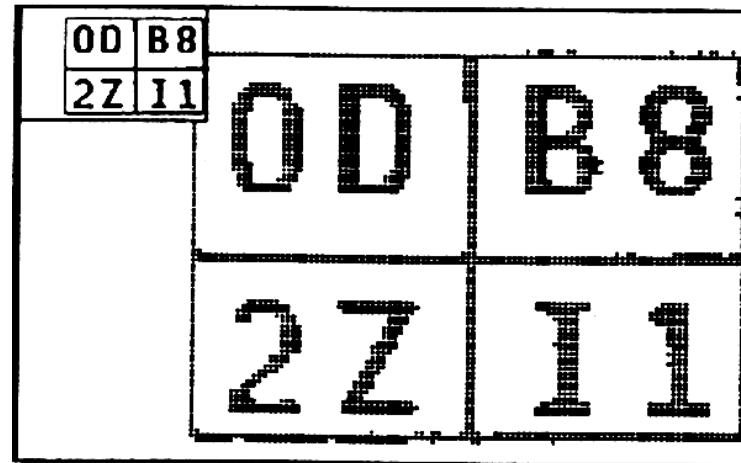




## چالش‌ها ...

### ○ تنوع بین دسته‌ای (Inter-Class Variability)

- نویسه‌های (کارکترهای) مشابه



- چهره‌های مشابه



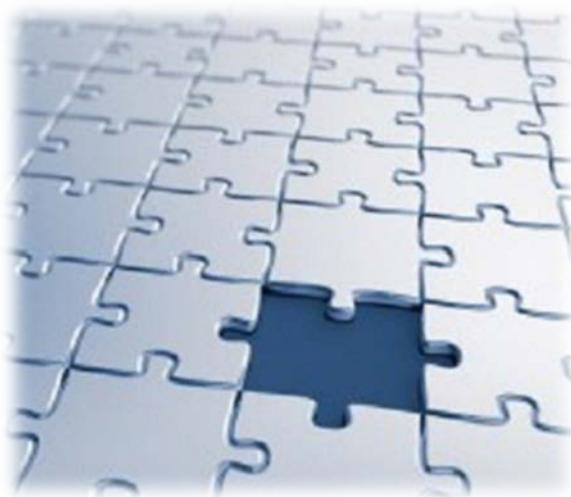
## چالش‌ها ...

### ○ داده‌ها/ویژگی‌های گم شده (Missing Data/Features)

- عدم امکام محاسبه مقدار برحی ویژگی‌ها (به دلیل نویز، ...)

• مثال

- دسته‌بندی ماهی: عدم امکان محاسبه طول یک ماهی به دلیل چسپیدن به یک ماهی دیگر
- پزشکی: عدم امکان محاسبه میزان اوره یک فرد در لحظه بررسی وضعیت وی





## چالش‌ها

### ◦ ویژگی‌های مطلوب

#### ◦ عدم حساس به تغییرات (Invariance)

- چرخاندن تصویر
- افزایش دامنه صدا
- ...

### ◦ دسته‌بندهای مطلوب

- تمایز بین دسته‌های مختلف (تنوع بین دسته‌ای)
- یکسان گرفتن تنوع درون دسته‌ای



# ارزیابی سیستم‌های یادگیری ماشین . . .



## ○ دقت/نرخ خطأ

- درصد بازشناسی درست/نادرست دسته‌ها
- نرخ قبول اشتباه/رد اشتباه

## ○ سرعت پردازش

## ○ هزینه طراحی و ساخت

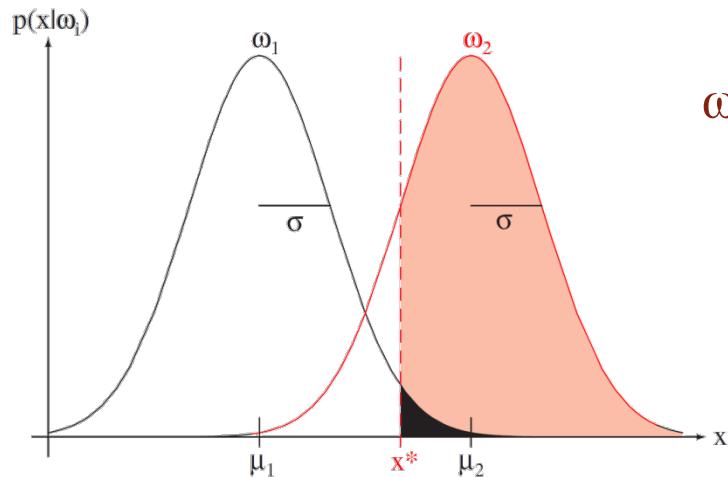
## ○ مقاوم بودن (Robustness)



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

$$p(x|\omega_1) \sim N(\mu_1, \sigma^2)$$

$$p(x|\omega_2) \sim N(\mu_2, \sigma^2)$$



## ○ دسته‌بندی دو حالته

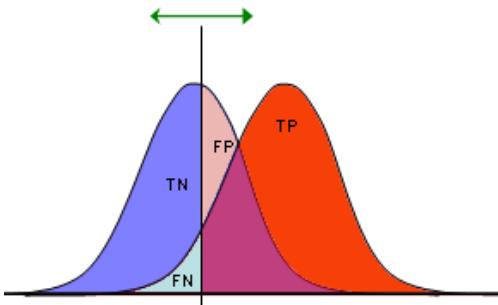
- تشخیص درست سیگنال برای میانگین  $\mu_1$  = دسته  $\omega_1$ 
  - دسته Negative (مثال: عدم تایید هویت گوینده)
- تشخیص درست سیگنال برای میانگین  $\mu_2$  = دسته  $\omega_2$ 
  - دسته Positive (مثال: تایید هویت گوینده)
- مرز تصمیم‌گیری = سطح آستانه  $x^*$

## ○ چهار نوع تصمیم

- |  |  |
|--|--|
| $P(x > x^*   x \in \omega_2)$                            | موفقیت یا مثبت صحیح (Hit=TruePositive) |
| $P(x < x^*   x \in \omega_2)$                            | خطا یا منفی غلط (Miss=FalseNegative)   |
| $P(x > x^*   x \in \omega_1)$ (FalseAlarm=FalsePositive) | هشدار نادرست یا مثبت غلط               |
| $P(x < x^*   x \in \omega_1)$                            | (CorrectRejection=TrueNegative)        |



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) ...



## چهار نوع تصمیم

- موفقیت یا مثبت صحیح (Hit=TruePositive)
- خطا یا منفی غلط (Miss=FalseNegative)
- هشدار نادرست یا مثبت غلط (FalseAlarm=FalsePositive)
- رد درست (CorrectRejection=TrueNegative)

مقدار خروجی سیستم		برای دو دسته	
نادرست (منفی) Negative	درست (مثبت) Positive	درست (w2) نادرست (w1)	مقدار واقعی
منفی غلط FN: False Negative	ثبت صحیح TP: True Positive	درست (w2)	
منفی صحیح TN: True Negative	ثبت غلط FP: False Positive	نادرست (w1)	



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

مقدار خروجی سیستم			
Negative	Positive	درست (فارسی)	مقدار واقعی
FN: False Negative - منفی غلط	TP: True Positive - مثبت صحیح	درست (فارسی)	
TN: True Negative - منفی صحیح	FP: False Positive - مثبت غلط	نادرست (سایر)	

## ○ معیارها

• درصد کل پاسخ‌های صحیح سیستم Accuracy

◦ برای وقتی که داده نامتوازن (Imbalanced Data) است، معیار خوبی نیست

◦ تعداد داده‌های یک دسته از دسته دیگر خیلی بیشتر/کمتر باشد

◦ برای وقتی که هزینه تصمیم سیستم برای دو دسته یکسان نیست، معیار خوبی نیست

◦ تشخیص خودکار افراد سرطانی

$$Accuracy = \frac{Total\ Corrects}{Total\ Data} = \frac{TP + TN}{TP + FN + FP + TN}$$

• (نرخ خطأ) درصد کل پاسخ‌های نادرست سیستم Error Rate

$$ErrorRate = \frac{Total\ Incorrects}{Total\ Data} = \frac{FN + FP}{TP + FN + FP + TN} = 1 - Accuracy$$



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

مقدار خروجی سیستم			
Negative	Positive	درست	مقدار واقعی
FN: False Negative - منفی غلط	TP: True Positive - مثبت صحیح	درست	
TN: True Negative - منفی صحیح	FP: False Positive - مثبت غلط	نادرست	

## ○ معیارها

• **Recall:** چند درصد از نمونه‌های درست، به درستی تشخیص داده شده‌اند

$$Recall = Sensitivity = TruePositiveRate(TPR) = \frac{Correct\ Things\ Found}{Correct\ Things\ Possible\ to\ Find} = \frac{TP}{Total\ Possitives\ =\ (TP + FN)}$$

تمامی نمونه‌های واقعی درست

• **Precision:** چند درصد از خروجی‌های درست تشخیص داده شده، واقعاً درست است

○ اگر فقط بکی از پاسخ‌های درست واقعی، درست تشخیص داده شده باشد و همه پاسخ‌های دیگر منفی باشند، مقدار این معیار ۱۰۰٪ است.

$$Precision = \frac{Correct\ Things\ Found}{Positive\ Things\ Found} = \frac{TP}{TP + FP}$$

• افزایش یکی از این معیارها، باعث کاهش دیگری می‌شود



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

مقدار خروجی سیستم			
Negative	Positive	درست	مقدار واقعی
FN: False Negative - منفی غلط	TP: True Positive - مثبت صحیح	درست	
TN: True Negative - منفی صحیح	FP: False Positive - مثبت غلط	نادرست	

## ○ معیارها

Recall و Precision و F-Measure •

- افزایش یکی از معیارهای Recall و Precision، باعث کاهش دیگری می‌شود
- میانگین هارمونیک دو عدد خیلی نزدیک به مقدار کمینه دو عدد است

$$F = \frac{1}{\alpha \cdot \frac{1}{Precision} + (1 - \alpha) \frac{1}{Recall}} = \frac{(\beta^2 + 1)Precision \cdot Recall}{\beta^2 Precision + Recall} \xrightarrow{\beta=1 \rightarrow \alpha=0.5} F = \frac{2Precision \cdot Recall}{Precision + Recall}$$

- ایراد: درصد تشخیص درست منفی‌ها (TN) در نظر گرفته نمی‌شود



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

مقدار خروجی سیستم	
Negative	Positive
FN: False Negative - منفی غلط	TP: True Positive - مثبت صحیح
TN: True Negative - منفی صحیح	FP: False Positive - مثبت غلط
<div style="display: flex; justify-content: space-around;"> <span>درست</span> <span>نادرست</span> </div> <span>مقدار واقعی</span>	

## ○ معیارها

• Sensitivity: چند درصد از نمونه‌های درست، به درستی تشخیص داده شده‌اند

$$Sensitivity = Recall = TruePositiveRate(TPR) = \frac{Correct\ Things\ Found}{Correct\ Things\ Possible\ to\ Find} = \frac{TP}{Total\ Possitives} = \frac{TP}{TP + FN}$$

• Specificity: چند درصد از نمونه‌های نادرست، به درستی تشخیص داده شده‌اند

○ معادل همان Sensitivity اما برای نمونه‌های نادرست

$$Specificity = TrueNegativeRate(TNR) = \frac{TN}{Total\ Negatives} = \frac{TN}{FP + TN}$$



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

مقدار خروجی سیستم			
Negative	Positive	درست	مقدار واقعی
FN: False Negative - منفی غلط	TP: True Positive - مثبت صحیح	درست	
TN: True Negative - منفی صحیح	FP: False Positive - مثبت غلط	نادرست	

## ● معیارها

● هشدار نادرست یا مثبت غلط (FalseAlarm=FalsePositive)

$$FalsePositiveRate = 1 - Specificity = \frac{FP}{Total\ Negatives} = \frac{FP}{TN + FP}$$

● موفقیت یا مثبت صحیح (Hit=TruePositive)

$$TruePositiveRate(TPR) = Recall = Sensitivity = \frac{TP}{Total\ Possitives} = \frac{TP}{TP + FN}$$

● افزایش یکی از این دو معیار، باعث کاهش دیگری می‌شود

● نمودار (Receiver Operating Characteristics) ROC



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) . . .

- ماتریس در هم ریختگی (Confusion Matrix)
- بیان می‌کند نمونه‌های هر دسته به عنوان چه دسته‌ای تشخیص داده شده‌اند

خروجی سیستم

- مثال: تشخیص ژانر موسیقی محلی ایرانی

سبک	بلوچی	بندری	خراسانی	کردی	لری	شیرازی	شمالی	ترکی
بلوچی	۷۶/۲	۱/۵	۱/۰۹	۴/۰۴	۱/۸	۱/۳	۶/۶	۷/۳
بندری	۷/۰۷	۷۵/۵	۰	۴/۶	۲/۳	۰	۶/۵	۳/۳
خراسانی	۴/۹	۱/۰۵	۸۲/۵	۲/۲	۱/۸	۰	۴/۳	۲/۸
کردی	۴/۴	۰	۱/۷	۷۴/۵	۳/۹	۰	۳/۹	۹/۶
لری	۱/۷	۱/۳	۱/۴	۴/۶	۸۵/۱	۰	۲/۴	۲/۹
شیرازی	۴/۴	۱/۶	۱/۴	۳/۱	۴/۲	۶۹/۲	۸/۸	۶/۹
شمالی	۴/۶	۰	۱/۹	۳/۶	۳/۶	۰	۷۸/۳	۶/۲
ترکی	۵/۰۷	۱/۰۳	۱/۲	۶	۴	۰	۵/۶	۷۶/۱

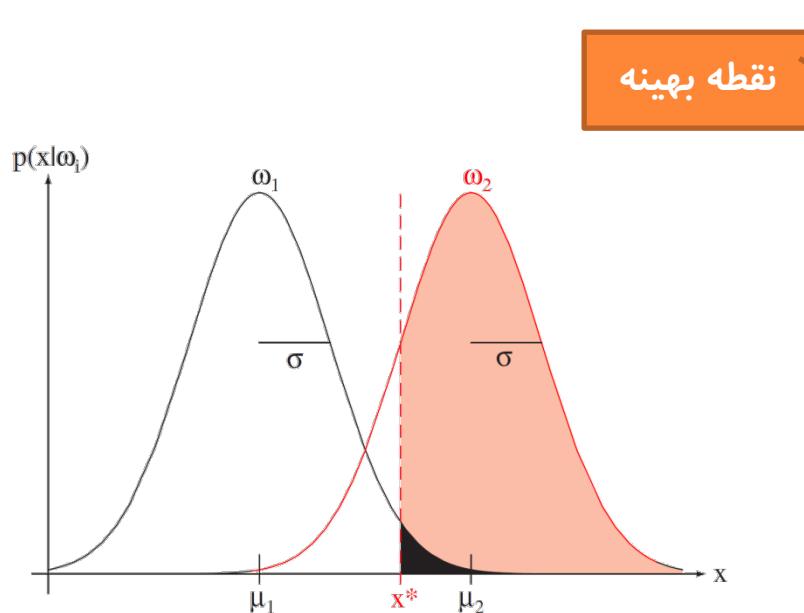
در ۳/۹٪ حالات موسیقی  
کردی به عنوان لری  
تشخیص داده می‌شود



# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) ...

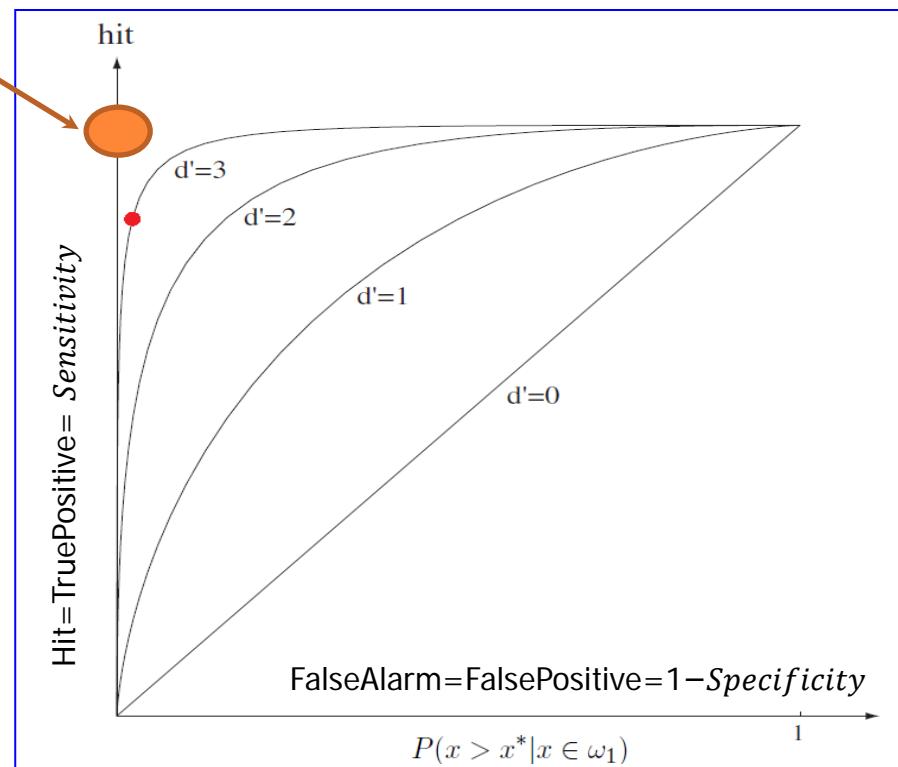
## نمودار (Receiver Operating Characteristics) ROC

- نمودار نرخ موفقیت (Hit) بر حسب نرخ هشدار نادرست (False Alarm)
- تفاوت مقدار  $x^*$  باعث = تغییر نرخ موفقیت (Hit) و نرخ هشدار نادرست (False Alarm)



$P(x > x^* | x \in \omega_2)$  Hit

$P(x > x^* | x \in \omega_1)$  False Alarm





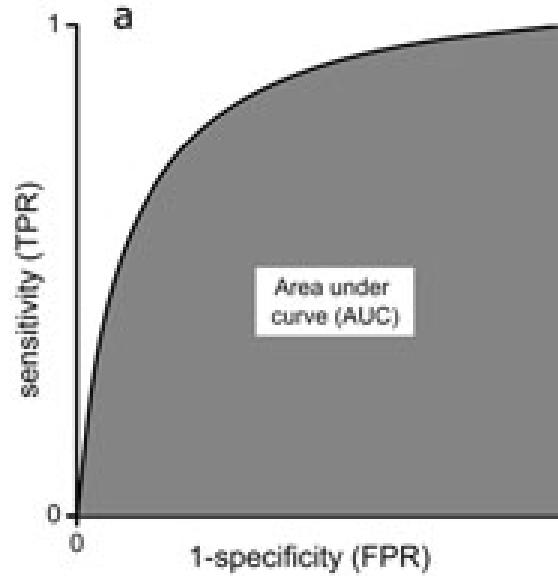
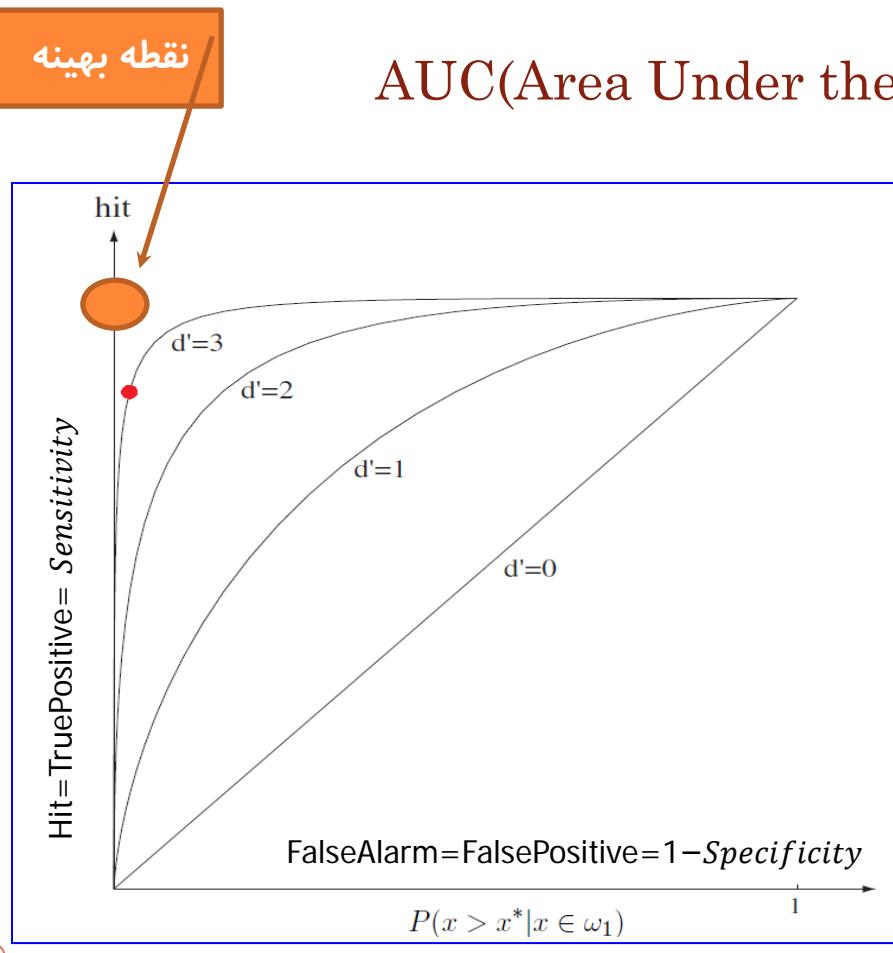
# ارزیابی سیستم‌های یادگیری ماشین (با ناظر) ...

## نمودار ROC (Receiver Operating Characteristics)

- تبدیل نمودار ROC به یک عدد

- محاسبه مساحت زیر منحنی (AUC)

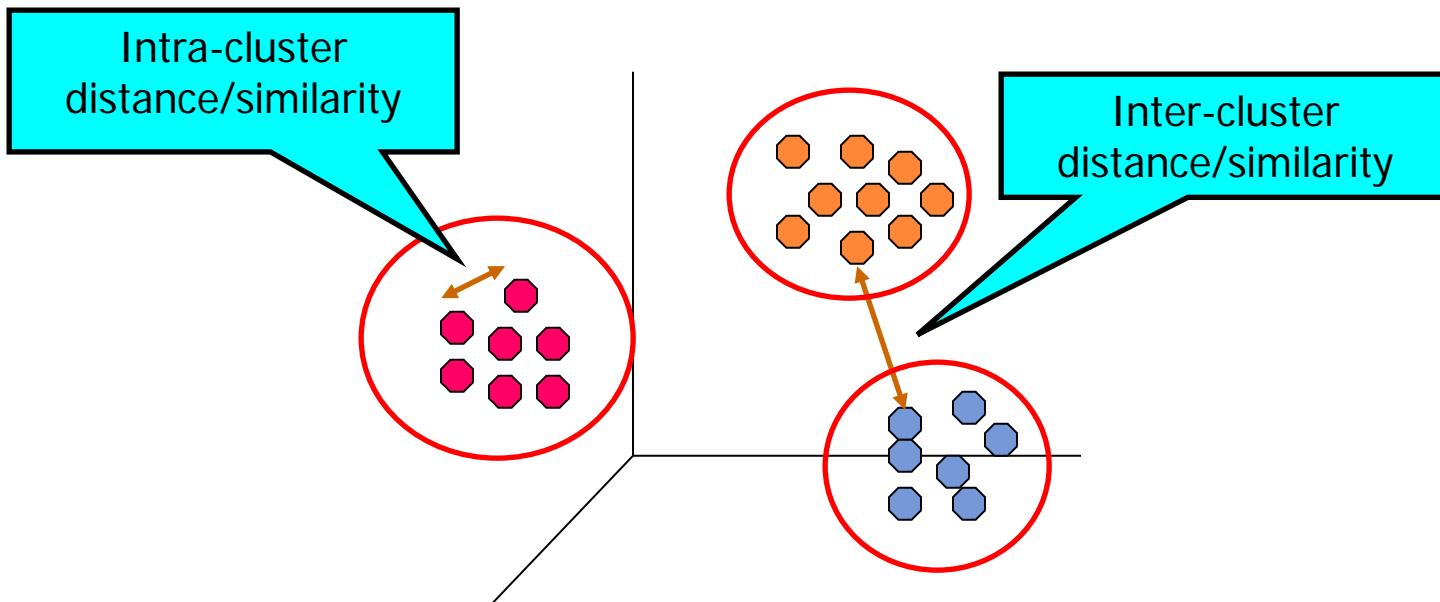
- هرچه مقدار AUC بیشتر باشد، بهتر است



# ارزیابی سیستم‌های یادگیری ماشین (بی‌ناظر) ...

## ○ معیارهای داخلی (Internal Criterion)

- داده‌های داخل هر خوشه بیشترین شباهت را به هم داشته باشند (Intra-Cluster Similarity)
- داده‌های بین خوشه‌ها کمترین شباهت را به هم داشته باشند (Inter Cluster Similarity)
- استفاده از معیارهای شباهت درون خوشه‌ای و بین خوشه‌ای





## ارزیابی سیستم‌های یادگیری ماشین (بی ناظر) ...

### ○ معیارهای خارجی (External Criterion)

- از یک مجموعه آزمون دارای برچسب استفاده می‌کند (Gold Standard)
  - دسته‌های (خوشه‌های) هر داده مشخص است
- ارزیابی بر اساس اینکه داده‌های هر دسته در یک خوشه قرار گرفته‌اند یا نه
- استفاده از آنتروپی یا معیارهای ارزیابی دسته‌بندی (مانند F-Measure)

### ○ معیارهای وابسته به کاربرد

- اگر هدف خوشه‌بندی، بهبود سرعت جستجو است، چقدر برای این منظور موفق بوده است
- پر هزینه است



# ارزیابی روش‌های بی ناظر: معیارهای داخلی . . .

## ◦ شباهت درون خوشه‌ای (Intra-Cluster Similarity)

- مطلوب آن است که داده‌های داخل هر خوشه شباهت بیشتری به هم داشته باشند
- بیانگر Cluster Cohesion
- تابع شباهت (Similarity) می‌تواند هر معیاری باشد، مانند شباهت کسینوسی
- به جای شباهت می‌توان از تابع فاصله استفاده کرد (شباهت بیشتر = فاصله کمتر)

$$Cohesion(C_k) = \sum_{x, y \in C_k} Similrity(x, y)$$

نمونه‌های متعلق به یک خوشه

## ◦ معیار SSE: Sum of Square Error

- جمع مربعات فاصله بین همه نمونه‌های یک خوشه با مرکز آن خوشه
- هرچه (فاصله SSE) کمتر باشد، بهتر است (شباهت بیشتر).

$$SSE(C_k) = \sum_{x \in C_k} (Dist(\mu_k, x))^2 \rightarrow SSE = \sum_{k=1}^K SSE(C_k) = \sum_{k=1}^K \sum_{i=1}^{|C_k|} (Dist(\mu_k, x_i))^2$$

مرکز خوشه

اندازه خوشه = تعداد نمونه‌ها

- فاصله می‌تواند هر معیاری باشد، از جمله فاصله اقلیدسی



## ارزیابی روش‌های بی ناظر: معیارهای داخلی . . .

### ○ شباهت بین خوشه‌ای (Inter Cluster Similarity)

- هرچه داده‌های بین خوشه‌ها شباهت کمتری داشته باشند، بهتر است.

### • بیانگر Cluster Separation

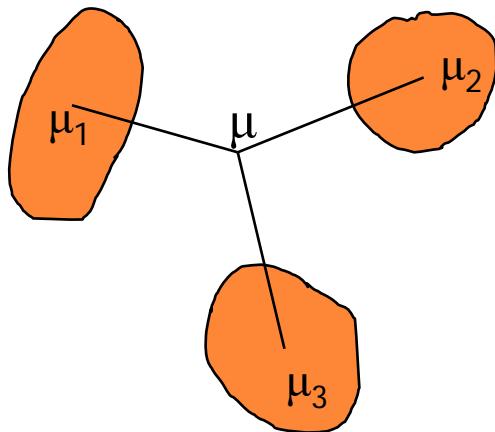
- تابع شباهت (Similarity) می‌تواند هر معیاری باشد، مانند شباهت کسینوسی

○ به جای شباهت می‌توان از تابع فاصله استفاده کرد (شباهت کمتر = فاصله بیشتر)

$$\text{Separation}(C_i, C_j) = \sum_{\substack{x \in C_i, \\ y \in C_j}} \text{Similarity}(x, y)$$

نمونه‌های متعلق به دو  
خوشه مختلف

# ارزیابی روش‌های بی ناظر: معیارهای داخلی . . .



## ○ معیار TSS: Total Sum of Squares

- محاسبه مرکز کلی خوشه‌ها (مرکز مرکزها!) =  $\mu$
- بیانگر فاصله همه نقاط از مرکز کلی (واریانس کل داده‌ها)

$$TSS = \sum_{i=1}^N (Dist(\mu, x_i))^2$$

## ○ معیار SSB: Sum of Square Between

- فاصله مراکز خوشه‌ها از مرکز کلی (بزرگتر باشد بهتر است)

$$SSB = \sum_{k=1}^K |C_k| (Dist(\mu_k, \mu))^2$$

## ○ داریم $TSS = SSE + SSB$

- برای یک مجموعه داده،  $TSS$  ثابت است
- اگر  $SSE$  (فاصله درون خوشه‌ای) افزایش یابد،  $SSB$  کاهش می‌یابد و برعکس



# ارزیابی روش‌های بی ناظر: معیارهای داخلی . . .

## ○ ضریب نیم رخ (Silhouette Coefficient)

• ترکیب شباهت درون خوشه‌ای (Separation) و بین خوشه‌ای (Cohesion)

• محاسبه برای یک نمونه داده مانند  $X_i$

○ گام ۱: محاسبه متوسط فاصله داده  $X_i$  از تمام داده‌های دیگر در خوشه خودش =  $a_i$

○ گام ۲: محاسبه متوسط فاصله داده  $X_i$  از تمام داده‌های دیگر در  $K-1$  خوشه دیگر (برای هر خوشه یک مقدار بدست می‌آید). کمترین مقدار بدست آمده از بین  $K-1$  متوسط فاصله محاسبه شده را انتخاب کن =  $b_i$

○ گام ۳: ضریب نیم رخ (Silhouette Coefficient)

• داریم  $-1 < s_i < 1$

○ مقدار منفی حالت نامناسب است (فاصله نمونه از سایر خوشه‌ها از خوشه خودش کمتر است)

○ حالت ایده آل: مقدار  $s_i = 1$  و  $a_i = b_i$

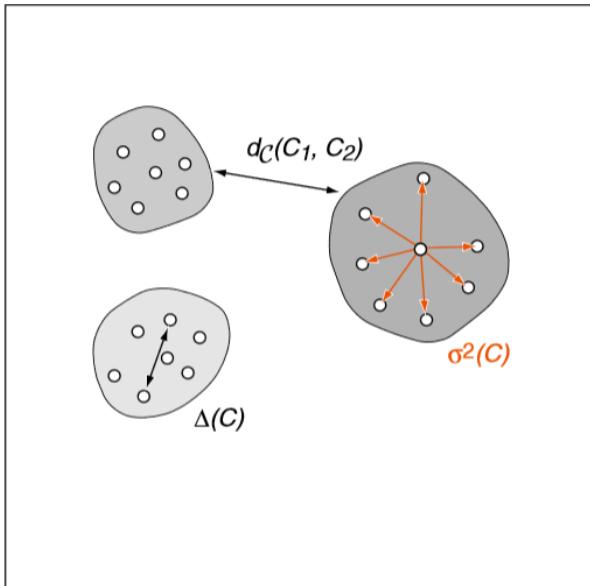
• بررسی مناسب بودن یک روش خوشه‌بندی: محاسبه متوسط  $s_i$ ‌ها برای کل داده‌ها

# ارزیابی روش‌های بی‌ناظر: معیارهای داخلی

## ● معیارهای دیگر

### ● همبستگی (Correlation)

- ترکیب شباهت درون خوشه‌ای (Cohesion) و بین خوشه‌ای (Separation)
- مقدار  $I(C)$  هرچقدر بیشتر باشد، بهتر است



$$I(C) = \frac{\min_{i \neq j} \{d_C(C_i, C_j)\}}{\max_{1 \leq l \leq k} \{\Delta(C_l)\}}$$



# ارزیابی روش‌های بی ناظر: معیارهای خارجی . . .

## ○ خالص بودن (Purity) خوشه‌ها

- دسته‌های واقعی هر نمونه داده مشخص است
- بعد از خوشه‌بندی، به هر خوشه یک برچسب بزن: برچسب دسته‌ای که بیشترین تعداد داده از آن دسته در این خوشه قرار گرفته است
  
- محاسبه درستی انتساب نمونه‌ها به خوشه‌ها
- محاسبه خالص بودن (Purity): شمارش تعداد نمونه‌های درست هر دسته، جمع زدن آنها با هم و تقسیم بر تعداد کل نمونه‌ها ( $N$ )

تعداد خوشه‌ها

$$\text{Purity}(\text{Cluster}, \text{Class}) = \frac{1}{N} \sum_{k=1}^K \max_j |\text{Cluster}_k \cap \text{Class}_j|$$

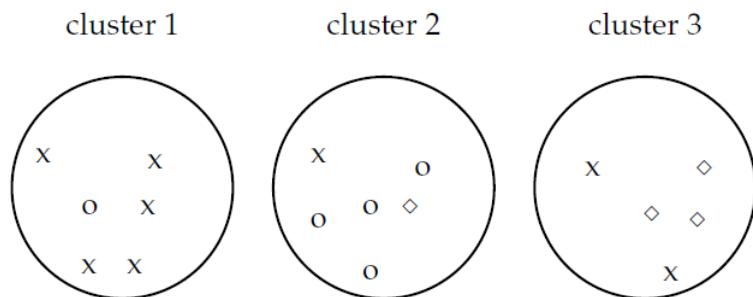
- مقدار خالص بودن بین ۱ (خوشه‌بندی بهینه) و ۰ (خوشه‌بندی بد) است
- وقتی تعداد خوشه‌ها زیاد باشد، مقدار خلوص افزایش می‌باید (هر نمونه در یک خوشه)



# ارزیابی روش‌های بی‌ناظر: معیارهای خارجی ...

## ○ خالص بودن (Purity) خوشه‌ها (مثال)

- ۱۷ نمونه داده داریم که متعلق به سه دسته (Class)  $\times$ ,  $\diamond$  و  $\circ$  هستند
- خوشه بندی در ۳ خوشه



$$\max_j |Cluster_1 \cap Class_j| = 5$$

$$\max_j |Cluster_1 \cap Class_j| = 4$$

$$\max_j |Cluster_1 \cap Class_j| = 3$$

$$Purity(Cluster, Class) = \frac{1}{N} \sum_{k=1}^K \max_j |Cluster_k \cap Class_j| = \frac{1}{17} (5 + 4 + 3) = 0.71$$



# ارزیابی روش‌های بی ناظر: معیارهای خارجی . . .

## ◦ اطلاعات متقابل نرمال شده (Normalized Mutual Information)

$$NMI(Cluster, Class) = \frac{I(Cluster, Class)}{[H(Cluster) + H(Class)]/2}$$

اطلاعات متقابل

آنتروپی

$$\begin{aligned} I(Cluster, Class) &= \sum_{k=1}^K \sum_{j=1}^J P(Cluster_k \cap Class_j) \log \frac{P(Cluster_k \cap Class_j)}{P(Cluster_k)P(Class_j)} \\ &= \sum_{k=1}^K \sum_{j=1}^J \frac{|Cluster_k \cap Class_j|}{N} \log \frac{|Cluster_k \cap Class_j| N}{|Cluster_k||Class_j|} \end{aligned}$$

تعداد خوشه ها      تعداد دسته ها

$$H(Cluster) = - \sum_{k=1}^K P(Cluster_k) \log P(Cluster_k) = - \sum_{k=1}^K \frac{|Cluster_k|}{N} \log \frac{|Cluster_k|}{N}$$



# ارزیابی روش‌های بی ناظر: معیارهای خارجی . . .

## ◦ اطلاعات متقابل نرمال شده (Normalized Mutual Information)

$$NMI(Cluster, Class) = \frac{I(Cluster, Class)}{[H(Cluster) + H(Class)]/2}$$

- مقدار اطلاعات متقابل I (قبل از نرمال کردن)
  - بیانگر افزایش میزان اطلاعات ما از دسته‌ها با دیدن خوشه‌ها
  - عددی بین صفر و یک
    - صفر = خوشه‌بندی تصادفی = دانستن خوشه کمکی به افزایش اطلاعات ما از دسته نمی‌کند
    - یک=هر خوشه دقیقاً بیانگر یک دسته باشد
  - افزایش تعداد خوشه‌ها (هر نمونه یک خوشه:  $N=K$ ) مقدار یک می‌ماند اما این خوشه‌بندی مناسب نیست
  - معيار ارزیابی باید به تعداد خوشه‌ها حساس باشد → نرمال کردن به مقادیر مخرج با افزایش تعداد خوشه‌ها مقدار  $H(Cluster)$  زیاد می‌شود → مقدار NMI کاهش می‌یابد
- مقدار NMI (بعد از نرمال کردن I): عددی بین صفر و یک



# ارزیابی روش‌های بی ناظر: معیارهای خارجی ...

## ○ آنتروپی

- محاسبه آنتروپی هر خوشه

احتمال تعلق یک عضو خوشه  $k$  به دسته  $j$

$$H(Cluster_k) = - \sum_{j=1}^J p_{jk} \log p_{jk}$$

- سپس، میانگین گیری وزن دار روی آنتروپی همه خوشه‌ها

$$H = \sum_{k=1}^K \frac{|Cluster_k|}{N} H(Cluster_k)$$



# ارزیابی روش‌های بی‌ناظر: معیارهای خارجی

## ○ معیار F-Measure و Rand Index

### • محاسبه با بدست آوردن

◦ TP: نمونه به خوش خودش (حاوی نمونه‌های مشابه) انتساب داده شده است

◦ TN: دو نمونه غیرمشابه به دو خوش مختلف انتساب داده شده است

◦ FP: نمونه به خوش دیگری (حاوی نمونه‌های غیرمشابه) انتساب داده شده است

◦ FN: دو نمونه مشابه به دو خوش مختلف انتساب داده شده است

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

### • محاسبه دقیقت انتساب

◦ بیانگر درصد (دقیقت) انتساب درست نمونه‌ها به خوش‌ها

### • محاسبه F-Measure

◦ ابتدا Precision و Recall محاسبه می‌شوند

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F_{\beta} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$



کلاس استاد شفیعی کدکنی - دانشگاه تهران

