مفاهیم و کاربردهای هوش مصنوعی در حسابداری

یادگیری ماشین، دانش تجربی



بهرام صادقی بی غم

عضو هیات علمی دانشگاه و مشاورصنعت

و تجارت در حوزه هوش مصنوعی و هوش تجاری

وبسایت رسمی www.sadeghibigham.ir/

چهار رویکرد توسعه هوش مصنوعی (یاداوری)

- رویکرد اول: الگوریتم (صرفا علم)
- رویکرد دوم: یادگیری ماشین (استخراج علم از تجربه)
- رویکرد سوم: یادگیری عمیق (صرفا تجربه)
- رویکرد چهارم: نوروسیمبولیک (استفاده الگوریتمی ازتجربه)

رویکرد دوم: یادگیری ماشین – تولید علم با تعدادی داده



- یافتن الگوها از میان داده ها
- یادگیری ماشین
- خوشه بندی
- طبقه بندی
- قواعد انجمنی، وابستگی ها و روابط
- پیش بینی

یادگیری ماشین چیست؟

- در رویکرد اول: الگوریتمی یا مبتنی بر قواعد (اگر آنگاه):
  - حل یک مسئله در کامپیوتر: نیاز به یک الگوریتم
  - وجود الگوریتم‌های مختلفی برای یک کار
  - کارآمدترین الگوریتم: پیچیدگی زمانی و فضایی کمتر

- برای برخی از وظایف الگوریتمی واضح و دقیق نداریم
  - مثال: پیش‌بینی رفتار مشتری
  - تشخیص ایمیل‌های اسپم از ایمیل‌های قانونی

- عصر "کلان داده"
- ما: تولیدکننده و مصرف‌کننده داده‌ها
- نیازمند محصولات و خدماتی تخصصی: درک نیازها و پیش‌بینی علایق
- وجود الگوی خاص در داده‌ها

یادگیری ماشین چیست؟

جبران کمبود دانش: با داده‌ها
گردآوری داده‌ها با ویژگیهای مشخص
مطلوب: یادگیری از این داده‌ها
وظیفه ما و ماشین (کامپیوتر): استخراج الگوریتم از روی داده های موجود

جایگاه یادگیری ماشین:

- ممکن است نتواند فرآیند را دقیق شناسایی کند، اما می‌تواند یک تقریب خوب و مفید بسازد.
- ممکن است این تقریب نتواند همه چیز را توضیح دهد، اما می‌تواند بخشی از داده‌ها را توضیح دهد.
- ممکن است شناسایی کل فرآیند امکان‌پذیر نباشد، اما می‌تواند الگوها یا نظم‌های خاصی را تشخیص دهد.

الگوها: به ما در درک فرآیند کمک کنند،
می توان از آنها برای پیش‌بینی استفاده کرد

یادگیری ماشین چیست؟

یادگیری ماشین: برنامه‌نویسی کامپیوترها برای بهینه‌سازی یک معیار عملکرد با استفاده از داده‌های نمونه یا تجربیات گذشته است.

وجود یک مدل تعریف‌شده با تعدادی پارامتر

یادگیری: اجرای یک برنامه کامپیوتری برای بهینه‌سازی پارامترهای مدل با استفاده از داده‌های آموزشی یا تجربیات گذشته

مدل ممکن است:

1. پیش‌بینی‌کننده (آینده)
2. توصیفی (گذشته و حال)
3. یا هر دو. باشد

یادگیری زمانی استفاده می‌شود که:

- تخصص انسانی وجود ندارد (پیمایش در مریخ)،
- انسان‌ها قادر به توضیح تخصص خود نیستند (تشخیص گفتار)
- راه‌حل در طول زمان تغییر می‌کند (مسیریابی در یک شبکه کامپیوتری)
- راه‌حل باید با موارد خاص تطبیق داده شود (بیومتریک کاربر)

یادگیری ماشین چیست؟

- **یادگیری مدل‌های عمومی از داده‌های نمونه‌های خاص**
- **داده‌ها ارزان و فراوان هستند؛ دانش گران و کمیاب**
- **ساختن مدلی که تقریب خوب و مفیدی از داده‌ها باشد.**

- نقش آمار در یادگیری ماشین: استنتاج از یک نمونه

- نقش علوم کامپیوتر در یادگیری ماشین : الگوریتم‌های کارآمد برای
  1. حل مسئله بهینه‌سازی
  2. نمایش و ارزیابی مدل برای استنتاج

زمینه های کاربرد

- خرده فروشی
- بانک‌های مالی
- تمام علوم و فنون دانشگاهی
- بخشهای وسیعی از صنعت و تجارت
- پزشکی
- مخابرات.
- مقادیر زیادی از داده‌ها در علم فیزیک، نجوم و زیست‌شناسی
- شبکه جهانی وب
- و تقریبا همه چیز

یادگیری ماشین چیست؟

یادگیری ماشین بخشی از هوش مصنوعی است و فقط یک مسئله پایگاه داده نیست
سیستمی که در یک محیط در حال تغییر قرار دارد باید توانایی یادگیری داشته باشد.

- ✓ اگر سیستم بتواند یاد بگیرد و با تغییرات سازگار شود، طراح سیستم نیازی به پیش‌بینی و ارائه راه‌حل برای همه موقعیت‌های ممکن ندارد.

یادگیری ماشین همچنین به ما کمک می‌کند تا برای بسیاری از مشکلات در بینایی، تشخیص گفتار و رباتیک راه‌حل‌هایی پیدا کنیم.

مثال: تشخیص چهره‌ها

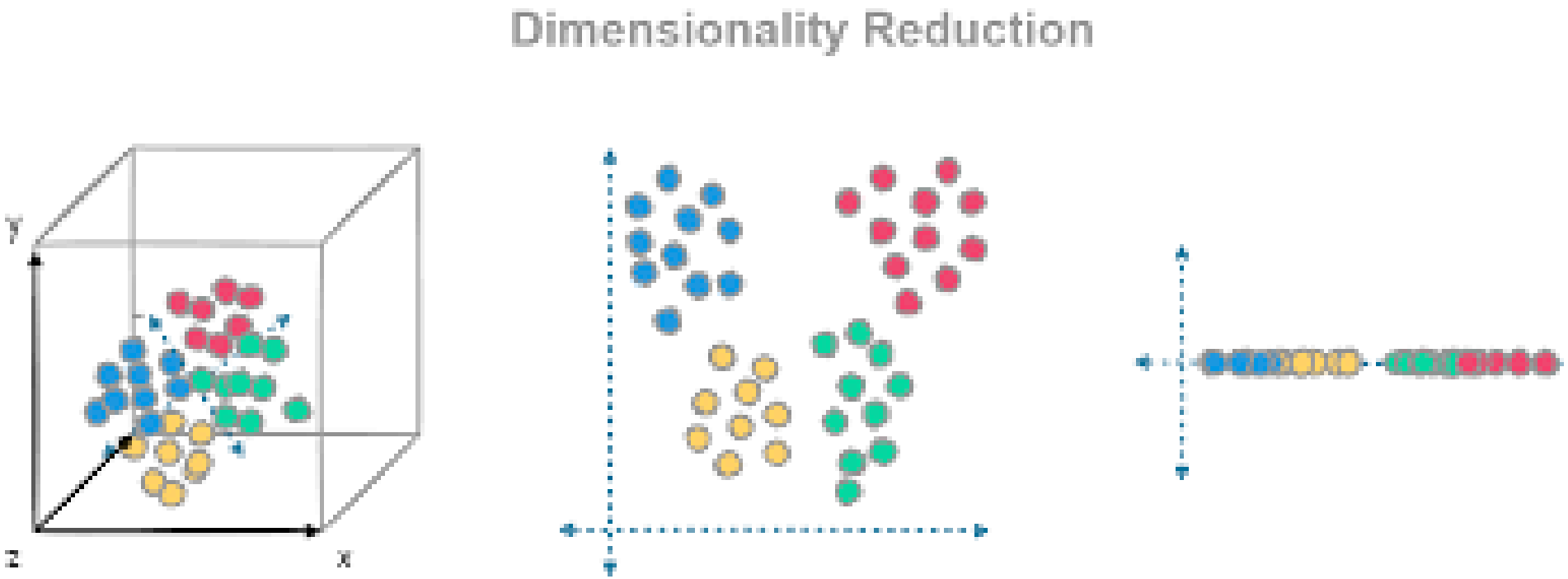
یک نمونه از تشخیص الگو

رویکردهای کلی

1. وابستگی و ارتباط (Association)
2. یادگیری با نظارت (Supervised Learning)
   - طبقه‌بندی (Classification)
   - رگرسیون (Regression)
3. یادگیری بدون نظارت (Unsupervised Learning)
4. یادگیری تقویتی (Reinforcement Learning)

کاهش ابعاد داده های ورودی

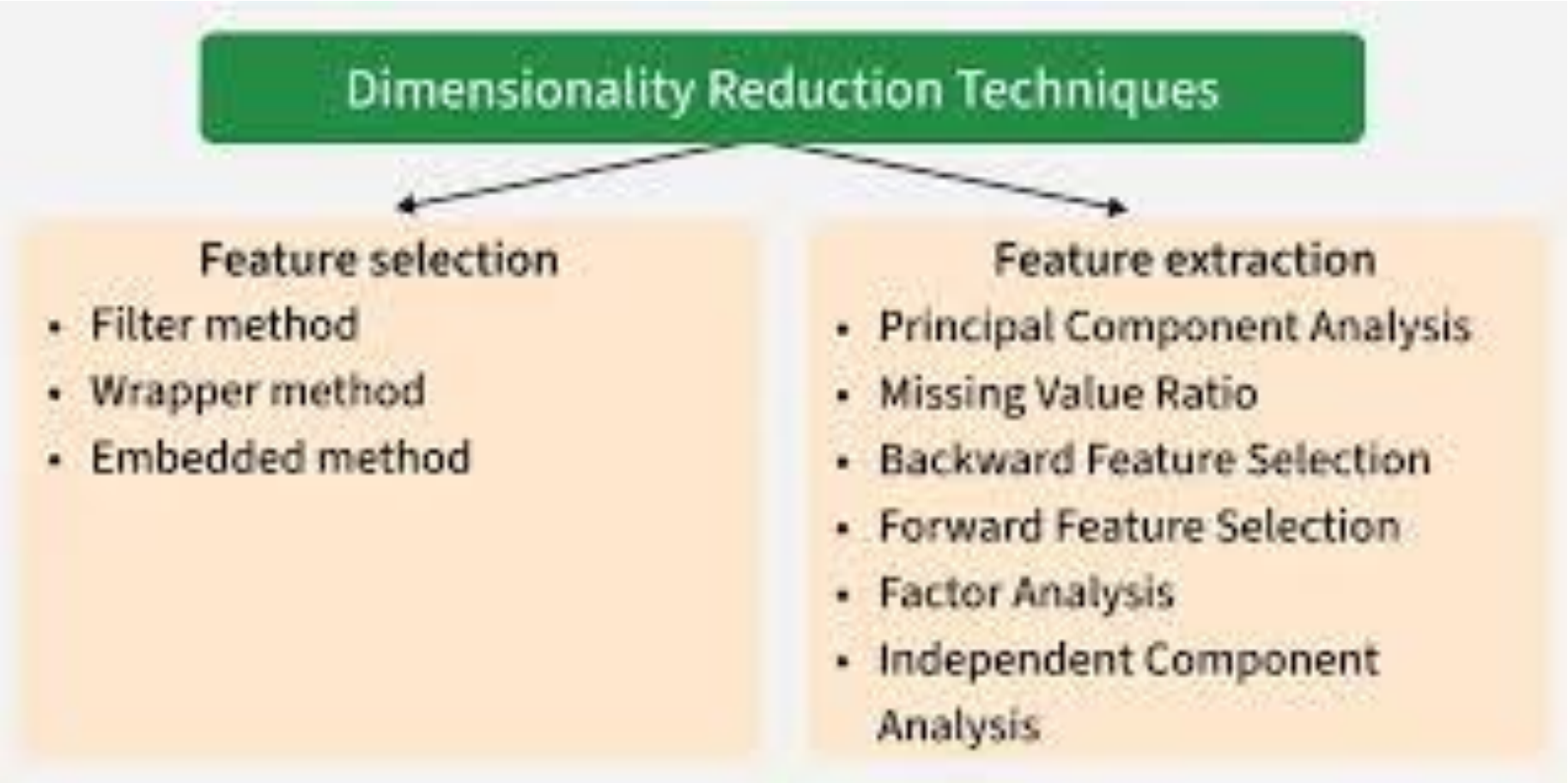
کاهش ابعاد روشی برای نمایش یک مجموعه داده معین با استفاده از تعداد کمتری از ویژگی‌ها (یعنی ابعاد) است، در حالی که همچنان ویژگی‌های معنادار داده‌های اصلی را ثبت می‌کند.



کاهش ابعاد داده های ورودی

Attribute	Domain	Type	Use in Network
A1 (Sex)	0, 1	Nominal	Candidate input
A2 (Age)	13.75 - 80.25	Continuous	Candidate input
A3 (Mean time at addresses)	0 - 28	Continuous	Candidate input
A4 (Home status)	1, 2, 3	Nominal	Candidate input
A5 (Current occupation)	1 - 14	Nominal	Candidate input
A6 (Current job status)	1 - 9	Nominal	Candidate input
A7 (Mean time with employers)	0 - 28.5	Continuous	Candidate input
A8 (Other investments)	0, 1	Nominal	Candidate input
A9 (Bank account)	0, 1	Nominal	Candidate input
A10 (Time with bank)	0 - 67	Continuous	Candidate input
A11 (Liability reference)	0, 1	Nominal	Candidate input
A12 (Account reference)	1, 2, 3	Nominal	Candidate input
A13 (Monthly housing expense)	0 - 2000	Continuous	Candidate input
A14 (Savings account balance)	1 - 100001	Continuous	Candidate input
Class (Reject / Accept)	0, 1	Nominal	Target

کاهش ابعاد داده های ورودی



یافتن ارتباطات و وابستگی ها

هدف: پیدا کردن ارتباط بین پدیده های مختلف در یک فضای مساله و میزان وابستگی و ارتباط آنها

association rule

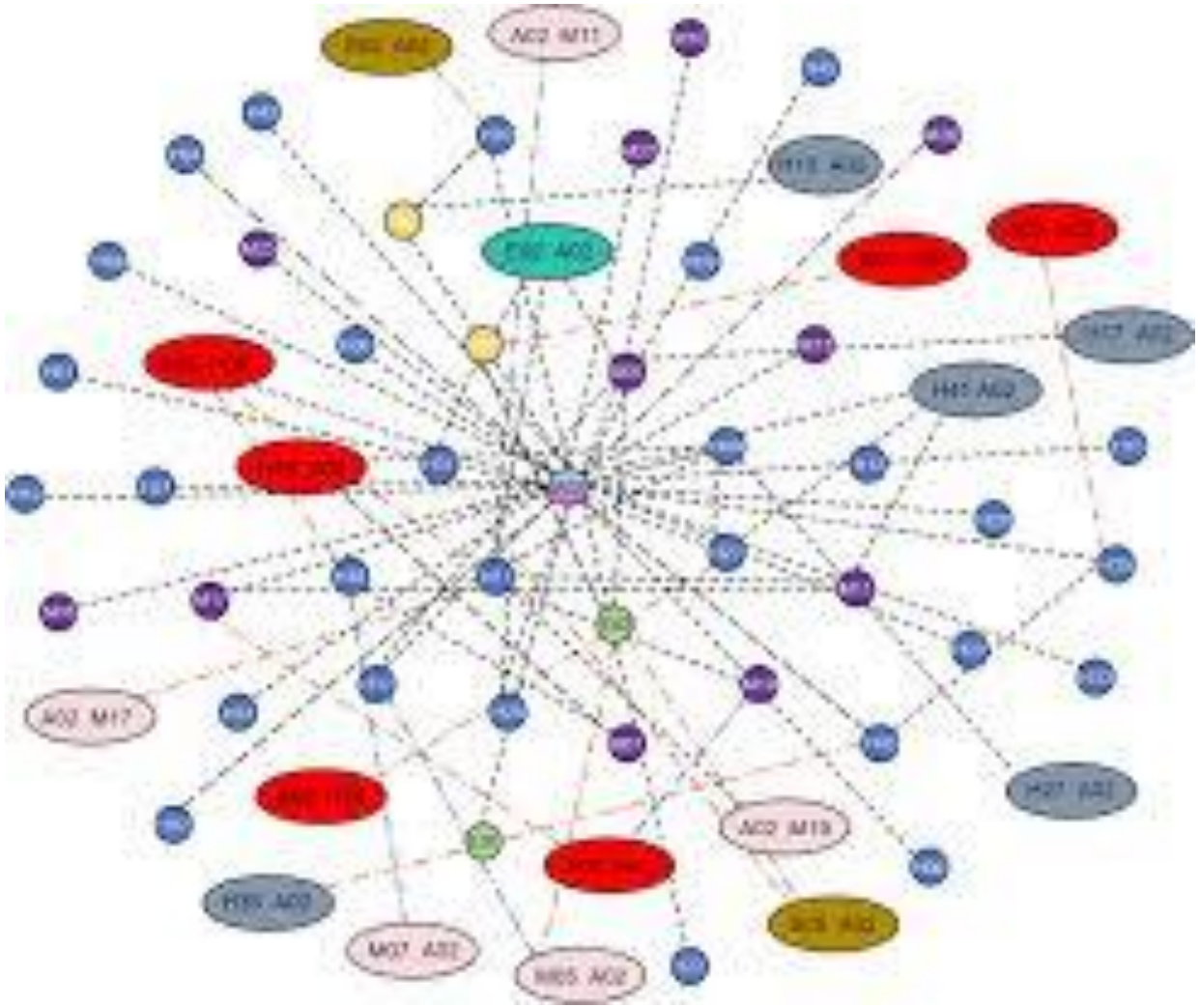
تحلیل سبد خرید:احتمال $P(Y \mid X)$ که کسی که X را می‌خرد، Y را نیز بخرد،

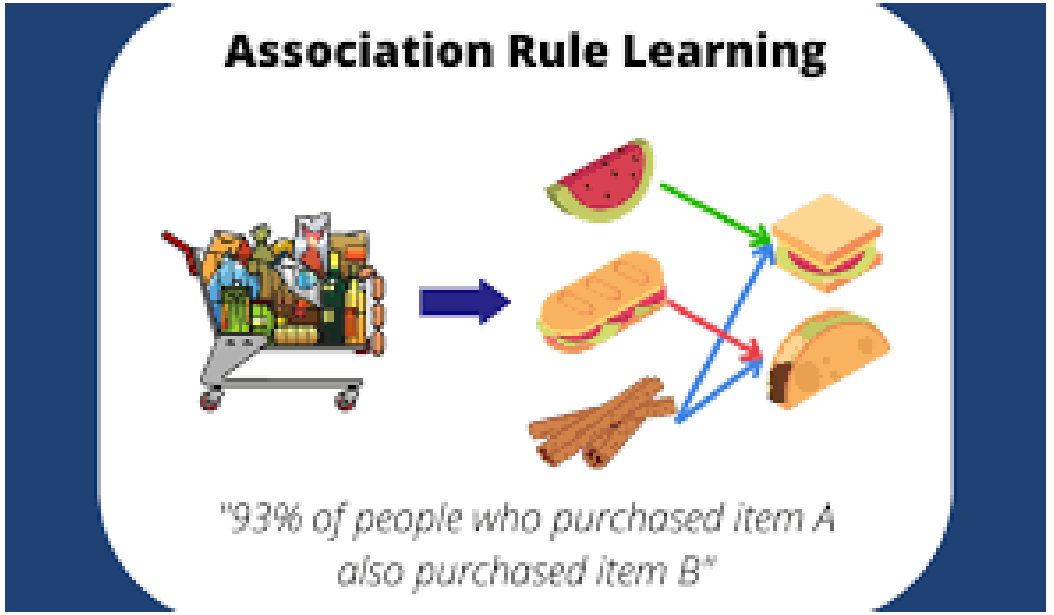
که در آن X و Y محصولات/خدمات هستند.

مثال: $P(Chips|Beer)=0.7$

با داشتن وابستگی ها می توان پیش بینی های لازم را کرد

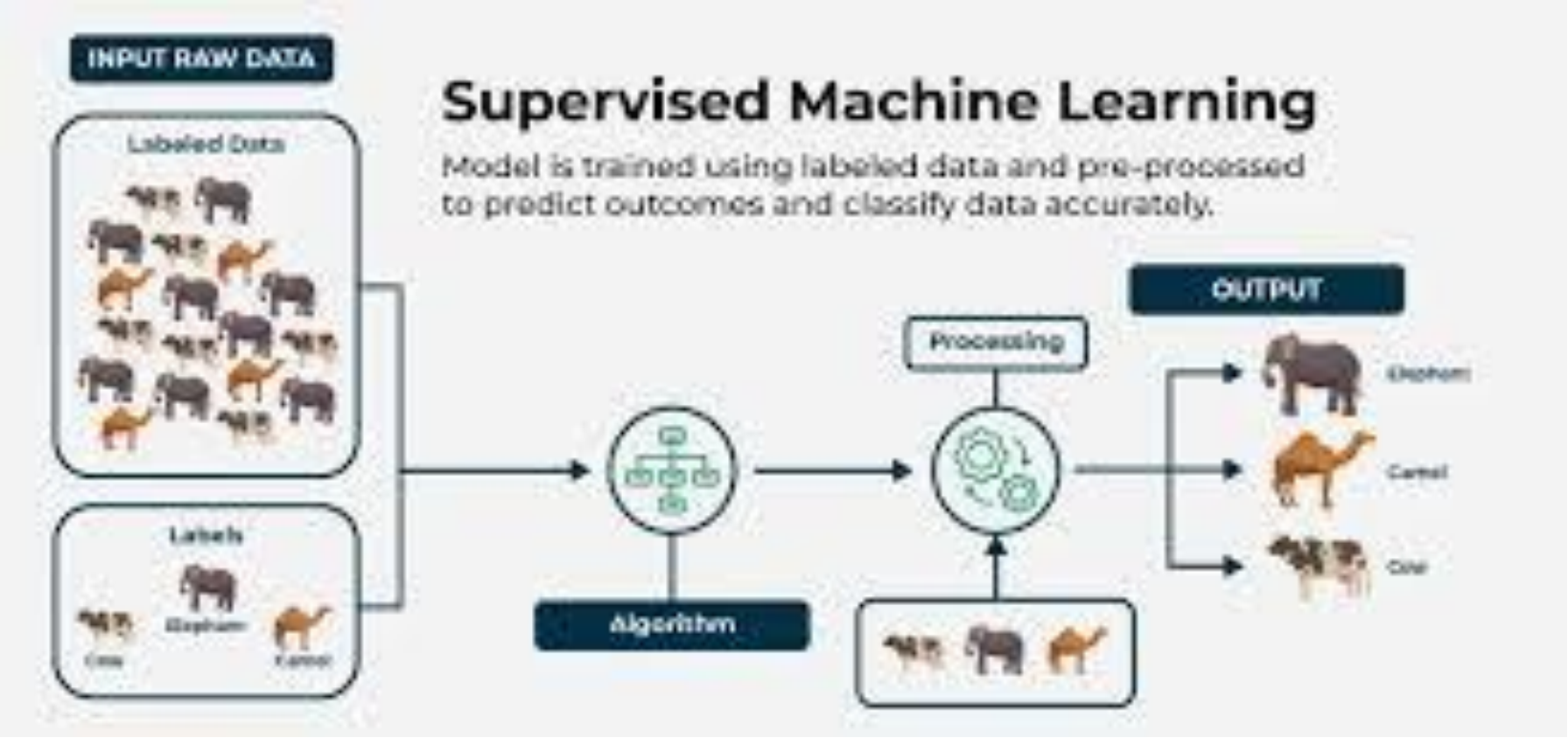
یافتن ارتباطات و وابستگی ها





یادگیری با نظارت

- یادگیری با نظارت یک تکنیک یادگیری ماشین است که از داده های آموزشی برچسب دار برای آموزش مدلهای هوش مصنوعی استفاده میکند تا رابطه الگوها را شناسایی کند.
- وظیفه مدل یادگیری نگاشت از ورودی به خروجی است.



انواع یادگیری با نظارت

انواع الگوریتم با نظارت

۱. رگرسیون (حالت پیوسته)

برخی الگوریتم‌ها: رگرسیون خطی، رگرسیون چندجمله‌ای.

.

۲. طبقه‌بندی (حالت گسسته)

برخی الگوریتم‌ها: درخت‌های تصمیم‌گیری، جنگل‌های تصادفی، ماشین‌های بردار پشتیبان

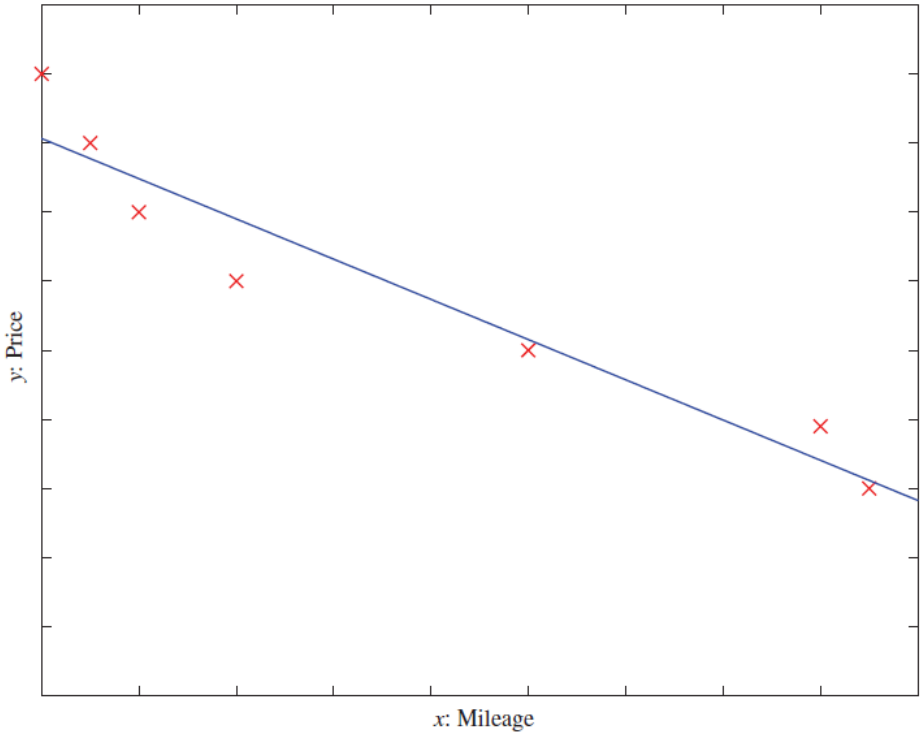
.

رگرسیون

رگرسیون: یک تکنیک یادگیری تحت نظارت است که برای پیش‌بینی مقادیر عددی پیوسته با یادگیری روابط بین متغیرهای ورودی (ویژگی‌ها) و یک متغیر خروجی (هدف) استفاده می‌شود.

- با متغیرهای خروجی با مقادیر حقیقی کار می‌کند.
- به شناسایی نقاط قوت و نوع روابط کمک می‌کند.
- از مدل‌های پیش‌بینی ساده و پیچیده پشتیبانی می‌کند.
- برای کارهایی مانند پیش‌بینی قیمت، پیش‌بینی روند و امتیازدهی ریسک استفاده می‌شود.

رگرسیون



مثال: پیش بینی قیمت یک ماشین دست دوم

x : ویژگی‌های ماشین

y : قیمت ماشین

$y = g(x \mid \theta)$

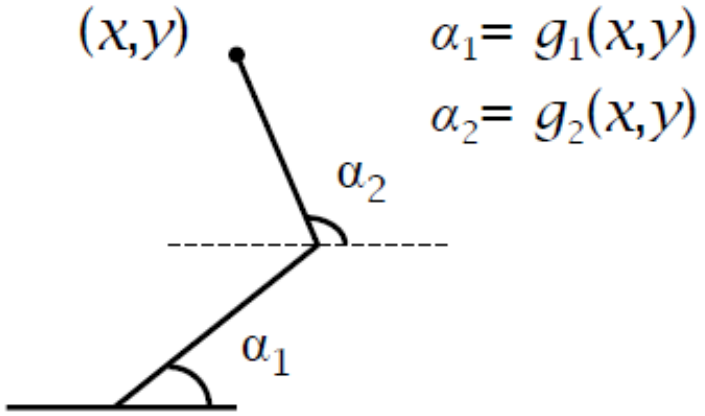
مدل g ()

پارامترهای θ

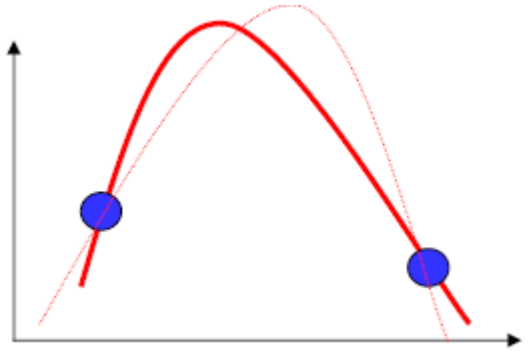
$$y = wx + w_0$$

رگرسیون: کاربرد

- ماشین های خودران: زاویه ای که فرمان باید بچرخد



- رباتهایی که بازو دارند و در کارخانه استفاده می شوند: تعیین زاویه ای که بازوی ربات باید بچرخد



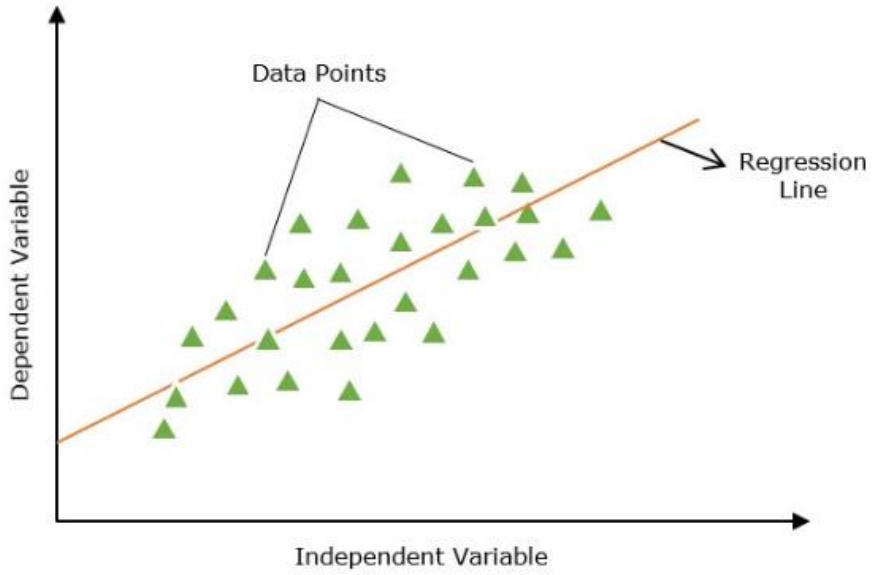
- طراحی سطح پاسخ

انواع رگرسیون

1. رگرسیون خطی ساده: رابطه بین یک متغیر مستقل و یک متغیر وابسته پیوسته را با برازش یک خط مستقیم که مجموع مربعات خطاها را به حداقل می‌رساند، مدل‌سازی می‌کند.

کاربرد: تخمین قیمت خانه فقط از روی اندازه آن
مزیت: به دلیل ساختار ریاضی ساده‌اش، قابلیت تفسیر بالایی دارد.
عیب: نمی‌تواند الگوهای داده‌های منحنی یا پیچیده را ثبت کند.

Square Feet (X)	House Price (Y)
1300	240
1500	320
1700	330
1830	295
1550	256
2350	409
1450	319



انواع رگرسیون

۲. رگرسیون خطی چندگانه: رگرسیون خطی چندگانه، رگرسیون خطی ساده را با ترکیب چندین متغیر مستقل برای پیش‌بینی یک نتیجه پیوسته، گسترش می‌دهد.

کاربرد: پیش‌بینی قیمت خانه با استفاده از عوامل متعددی مانند اندازه، موقعیت مکانی، قدمت و تعداد اتاق‌ها

مزیت: تأثیر ترکیبی بسیاری از عوامل را به طور همزمان در نظر می‌گیرد.

عیب: عملکرد در حضور هم‌خطی چندگانه (ویژگی‌هایی که همبستگی بالایی با یکدیگر دارند) کاهش می‌یابد.

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.8
162597.7	151377.6	443898.5	California	191792.1
153441.5	101145.6	407934.5	Florida	191050.4
144372.4	118671.9	383199.6	New York	182902
142107.3	91391.77	366168.4	Florida	166187.9
131876.9	99814.71	362861.4	New York	156991.1
134615.5	147198.9	127716.8	California	156122.5
130298.1	145530.1	323876.7	Florida	155752.6
120542.5	148719	311613.3	New York	152211.8

$$Y = w_0 + w_1 X_1 + w_2 X_2 + w_2 X_2$$

- X_1 represents R&D Spend,
- X_2 represents Administration, and
- X_3 represents Marketing Spend.

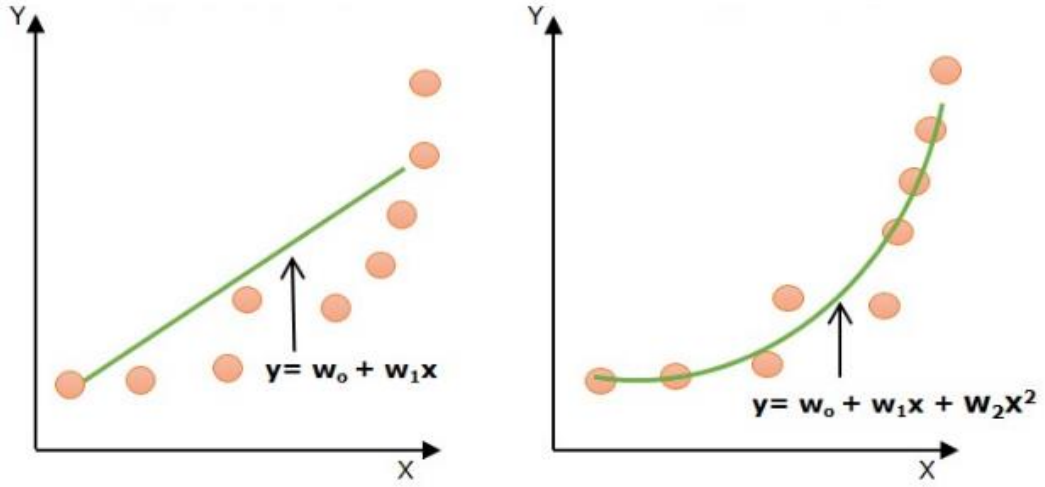
انواع رگرسیون

۳. رگرسیون چندجمله‌ای: رگرسیون چندجمله‌ای با تبدیل ویژگی‌های ورودی به عبارات چندجمله‌ای درجه بالاتر روابط غیرخطی را مدل‌سازی می‌کند.

کاربرد: مدل‌سازی روندهای رشد منحنی مانند افزایش جمعیت یا تغییرات دما

مزیت: روابط غیرخطی را بدون تغییر به الگوریتم‌های غیرخطی به طور مؤثر در نظر می‌گیرد.

عیب: چندجمله‌ای‌های درجه بالاتر ممکن است منجر به بیش‌برازش و پیش‌بینی‌های ناپایدار شوند.



.

$$y = w_0 + w_1x + w_2x^2 + w_3x^3 + \ldots + w_nx^n + \epsilon$$

طبقه‌بندی

- طبقه‌بندی : فرآیند پیش‌بینی کلاس یا دسته از داده های داده شده

طبقه‌بندی در یادگیری ماشین: یک تکنیک یادگیری نظارت شده است که در آن یک الگوریتم با داده‌های برچسب‌گذاری شده آموزش داده می‌شود تا دسته داده‌های جدید را پیش‌بینی کند.

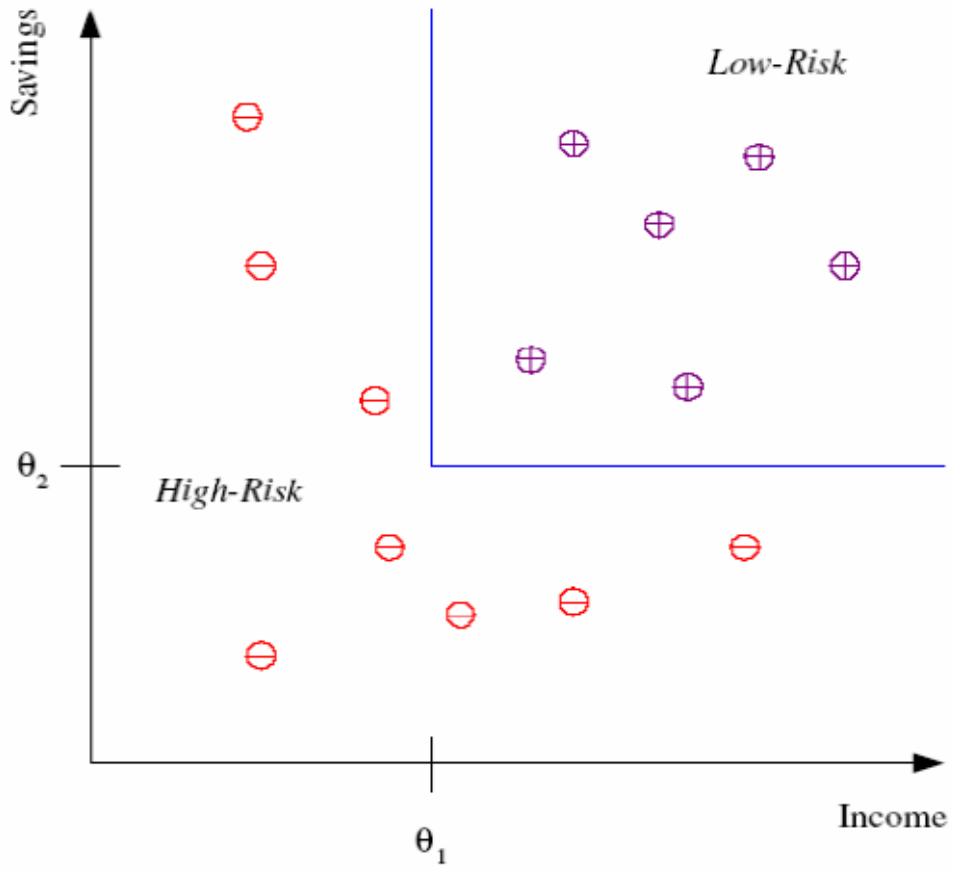
از نظر ریاضی، طبقه‌بندی وظیفه تقریب یک تابع نگاشت **(f)** از متغیرهای ورودی **(X)** به متغیرهای خروجی **(Y)** است. این اساساً متعلق به یادگیری ماشین نظارت شده است که در آن اهداف نیز همراه با مجموعه داده‌های ورودی ارائه می‌شوند.

پیاده‌سازی: ابتدا باید طبقه‌بندی‌کننده را آموزش دهیم. پس از آموزش موفقیت‌آمیز طبقه‌بندی‌کننده، می‌توان از آن برای تشخیص یک نمونه داده ناشناخته استفاده کرد.

طبقه بندی: مثال

مثال: امتیازدهی اعتباری

تمایز قائل شدن بین مشتریان کم‌ریسک و پرخطر از نظر درآمد و پس‌انداز آنها



مشخص‌کننده:

اگر درآمد > θ1 و پس‌انداز> θ2 باشد، آنگاه کم‌ریسک در غیر این صورت پرخطر

طبقه بندی: کاربردها

❖ تشخیص الگو

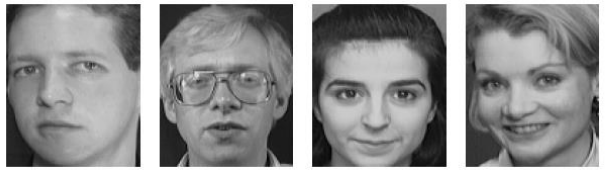
- تشخیص نوری کاراکتر: سبک‌های مختلف دست‌خط. یک کلمه، دنباله ای از کاراکترها است و کاراکترهای متوالی مستقل نیستند، بلکه توسط کلمات زبان محدود می شوند. این مزیت را دارد که از دانش زبان یک کاراکتر را تشخیص دهیم، هنوز [illegible] را بخوانیم.

- تشخیص چهره: ژست، نورپردازی، انسداد (عینک، ریش)، آرایش، مدل مو

Training examples of a person



Test images



طبقه بندی: کاربردها

❖ تشخیص الگو



- تشخیص پزشکی: از علائم تا بیماری‌ها

- تشخیص گفتار: وابستگی زمانی.
  - ° استفاده از فرهنگ لغت یا نحو زبان.
  - ° ادغام حسگر: ترکیب چندین روش؛ به عنوان مثال، بصری (تصویر لب) و صوتی برای گفتار



- ..

طبقه بندی: کاربردها

- استخراج دانش

یادگیری یک قانون از داده‌ها، استخراج دانش را نیز امکان‌پذیر می‌سازد.
این قانون، مدل ساده‌ای است که داده‌ها را توضیح می‌دهد و با نگاه به این مدل، توضیحی در مورد فرآیند زیربنایی داده‌ها خواهیم داشت.

- فشرده سازی

با برازش یک قانون به داده‌ها، توضیحی ساده‌تر از خود داده‌ها به دست می‌آوریم که به حافظه کمتری برای ذخیره و محاسبات کمتری برای پردازش نیاز دارد.

- تشخیص داده‌های پرت

یافتن مواردی است که از قانون پیروی نمی‌کنند و استثنا هستند. در این حالت، پس از یادگیری قانون، ما به خود قانون علاقه‌ای نداریم، بلکه به استثنائاتی که توسط قانون پوشش داده نمی‌شوند، علاقه‌مندیم که ممکن است نشان‌دهنده ناهنجاری‌هایی باشد که نیاز به توجه دارند

الگوریتم‌های طبقه‌بندی

هدف: یادگیری یک تابع نگاشت **(f)** بین ویژگی های ورودی **(X)** و متغیرهای هدف **(Y)** است.
این نگاشت اغلب با یک مرز تصمیم گیری نمایش داده می شود.

الگوریتم‌های مهم طبقه‌بندی یادگیری ماشین:

- **Logistic Regression**
- **K-Nearest Neighbors (KNN)**
- **Support Vector Machine (SVM)**
- **Decision Tree**
- **Naive Bayes**
- **Random Forest**

K-Nearest Neighbors (KNN)

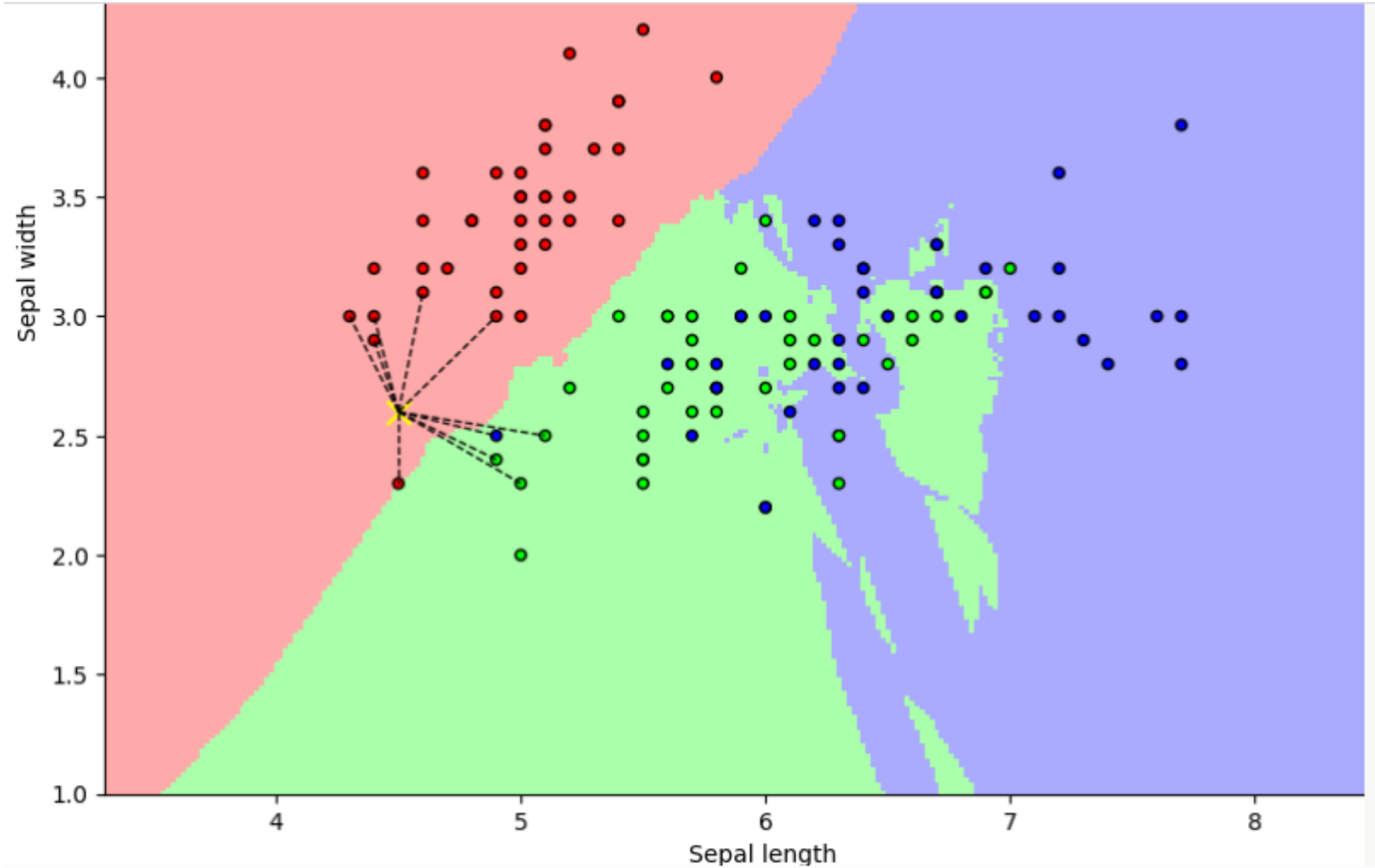
الگوریتم k –نزدیک‌ترین همسایه (KNN) : الگوریتم یادگیری ماشینی تحت نظارت است

ایده اصلی: یافتن k–نزدیک‌ترین نقاط داده به یک نقطه داده آزمایشی داده شده و استفاده از این نزدیکترین همسایه‌ها برای پیش‌بینی است.
k: یک پارامتر است که باید تنظیم شود و نشان دهنده تعداد همسایه‌هایی است که باید در نظر گرفته شوند.

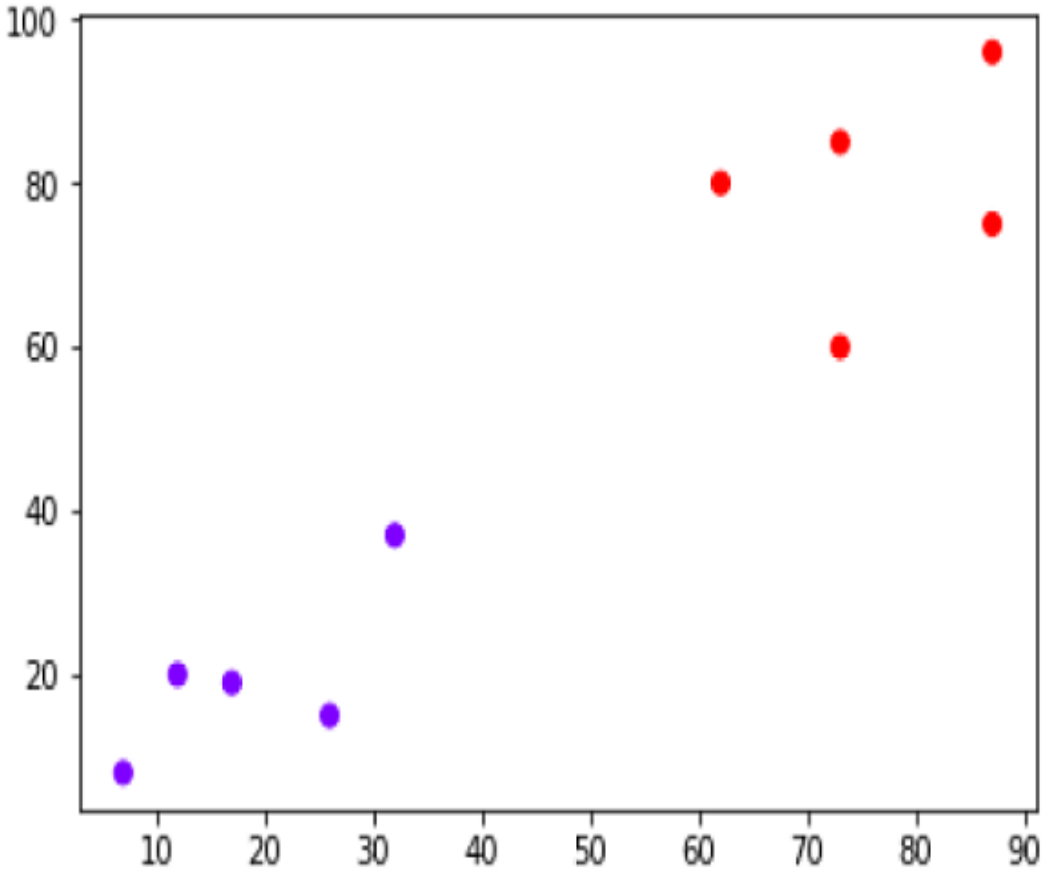
- **الگوریتم KNN در مسائل طبقه‌بندی: نقطه داده آزمایشی را به کلاسی اختصاص می‌دهد که بیشترین فراوانی را در بین k-نزدیک‌ترین همسایه‌ها دارد**
- **الگوریتم KNN در مسائل رگرسیون: میانگین مقادیر k-نزدیک‌ترین همسایه‌ها را به نقطه داده آزمایشی اختصاص می‌دهد.**

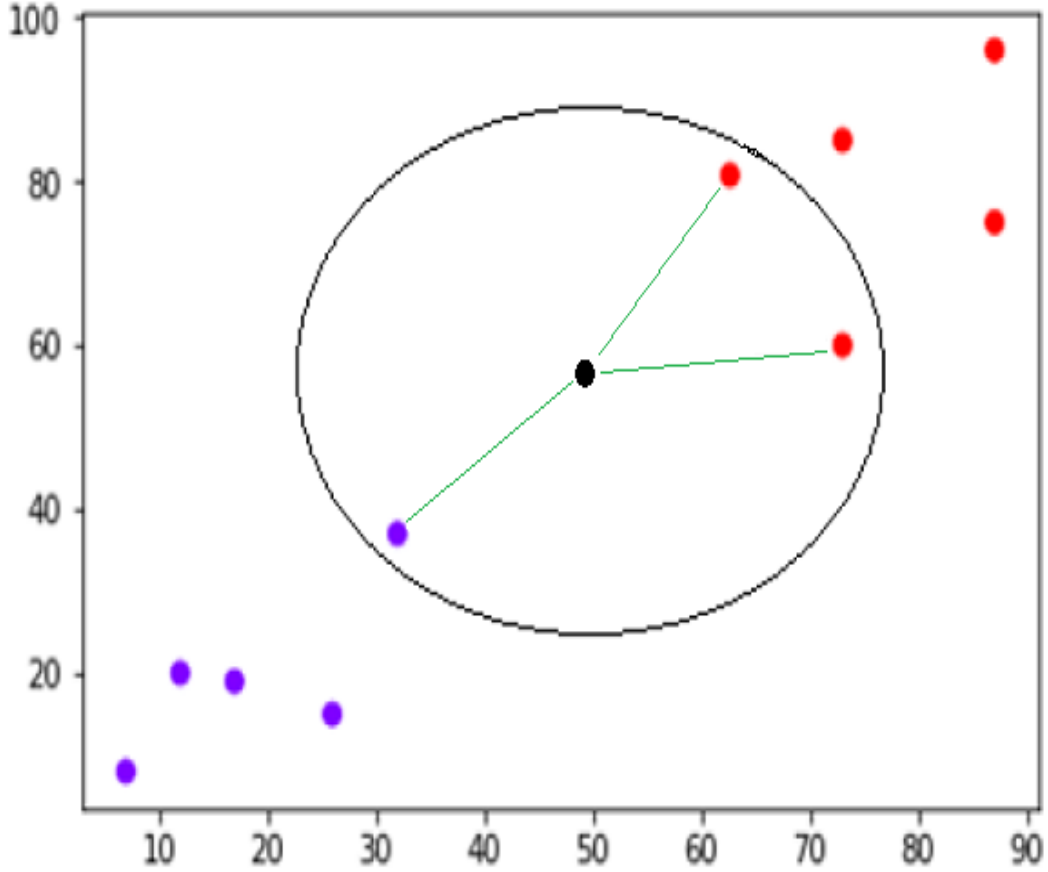
معیار فاصله: فاصله اقلیدسی، فاصله منهتن و غیره (متر های مختلف برای شباهت و اختلاف)

K-Nearest Neighbors (KNN)



کاربردها


100
80
60
40
20
10
20
30
40
50
60
70
80
90


100
80
60
40
20
10
20
30
40
50
60
70
80
90

درخت تصمیم

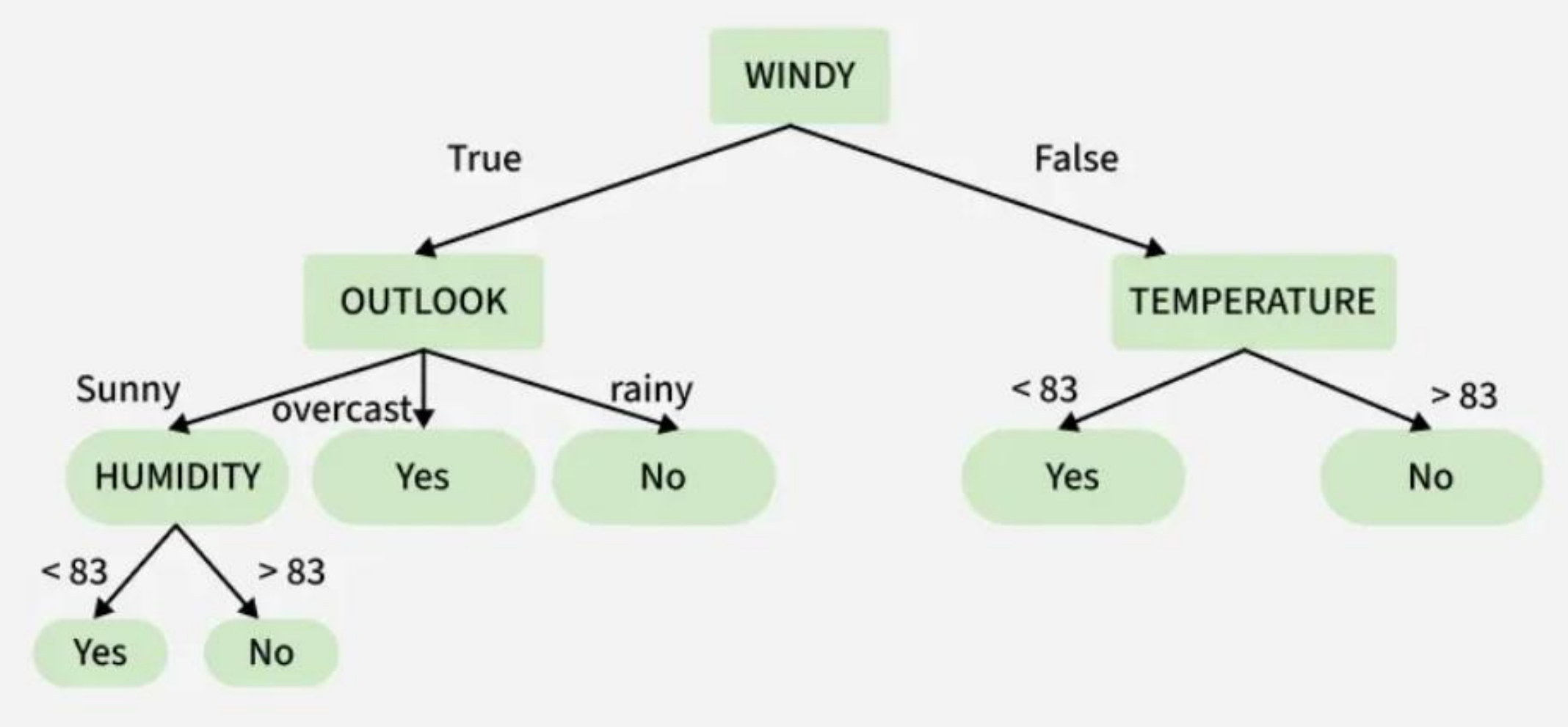
برای طبقه‌بندی یا پیش‌بینی نتایج بر اساس مجموعه‌ای از قوانین
تقسیم داده‌ها به زیرمجموعه‌هایی بر اساس مقادیر ویژگی‌های ورودی
انتخاب بهترین ویژگی برای تقسیم داده‌ها در هر گره

- بیشترین بهره اطلاعاتی
- بیشترین کاهش آنتروپی

تقسیم بازگشتی داده‌ها تا این که داده‌های هر زیرمجموعه به یک کلاس تعلق داشته باشند
درخت حاصل مجموعه‌ای از قوانین تصمیم‌گیری است
استفاده از درخت حاصل برای پیش‌بینی یا طبقه‌بندی داده‌های جدید

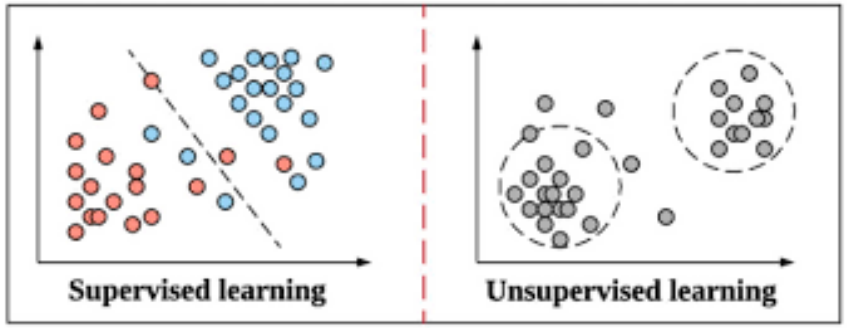
دو نوع الگوریتم درخت تصمیم:
۱. درخت طبقه‌بندی
۲. درخت رگرسیون

Decision Tree

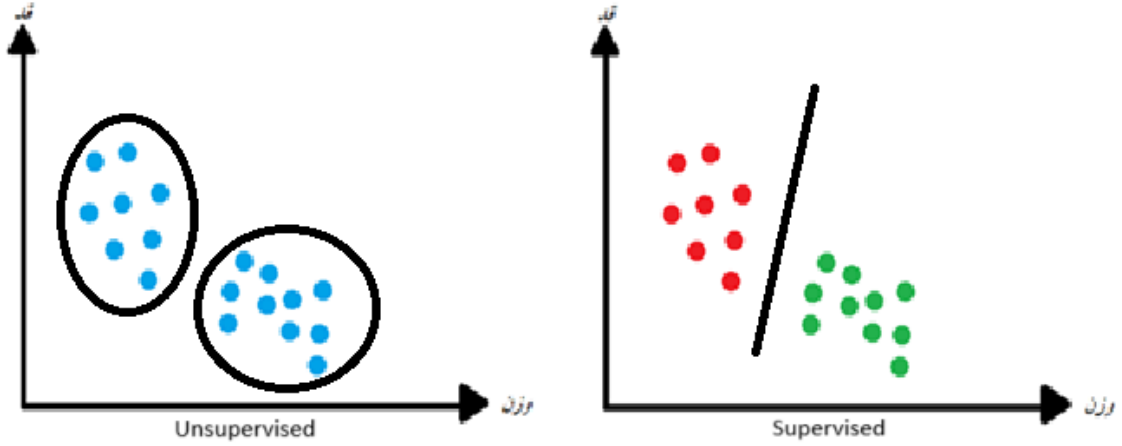


یادگیری بدون نظارت

در یادگیری بدون نظارت، ناظری وجود ندارد و ما فقط داده‌های ورودی بدون برچسب داریم.
یک رویکرد یا نوع یادگیری ماشین از الگوریتم‌های یادگیری ماشین برای یافتن الگوها یا ساختارهای پنهان درون داده‌ها بدون نظارت استفاده می‌کند.
رویکردهای یادگیری ماشین بدون نظارت: ارتباط، خوشه‌بندی و کاهش ابعاد.



کاربرد: در همه علوم و فنون و اصلی ترین موضع یادگیری ماشین



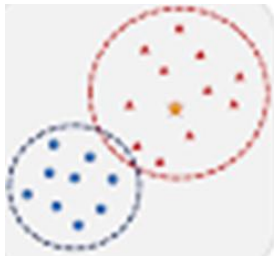
یادگیری بدون نظارت

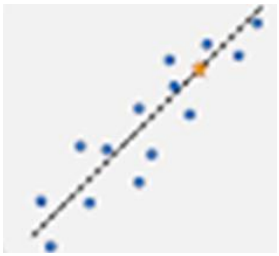
مقایسه

Supervised

Data:
Labeled examples
(input , desired output)

Tasks:
classification
pattern recognition
regression

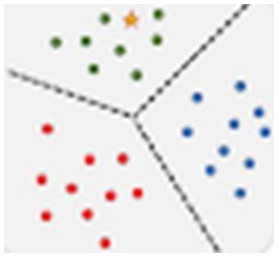




Unsupervised

Data:
Unlabeled examples
(different realizations of the input)

Tasks:
clustering



الگوریتم K-Means Clustering

مرحله ۱ – ابتدا باید تعداد خوشه‌ها، **K**، را که باید توسط این الگوریتم تولید شوند، مشخص کنیم.

مرحله ۲ – سپس، به طور تصادفی **K** نقطه داده را انتخاب کرده و هر نقطه داده را به یک خوشه اختصاص دهید. به عبارت ساده، داده‌ها را بر اساس تعداد نقاط داده طبقه‌بندی کنید.

مرحله ۳ – حالا مراکز خوشه‌ها محاسبه می‌شوند.

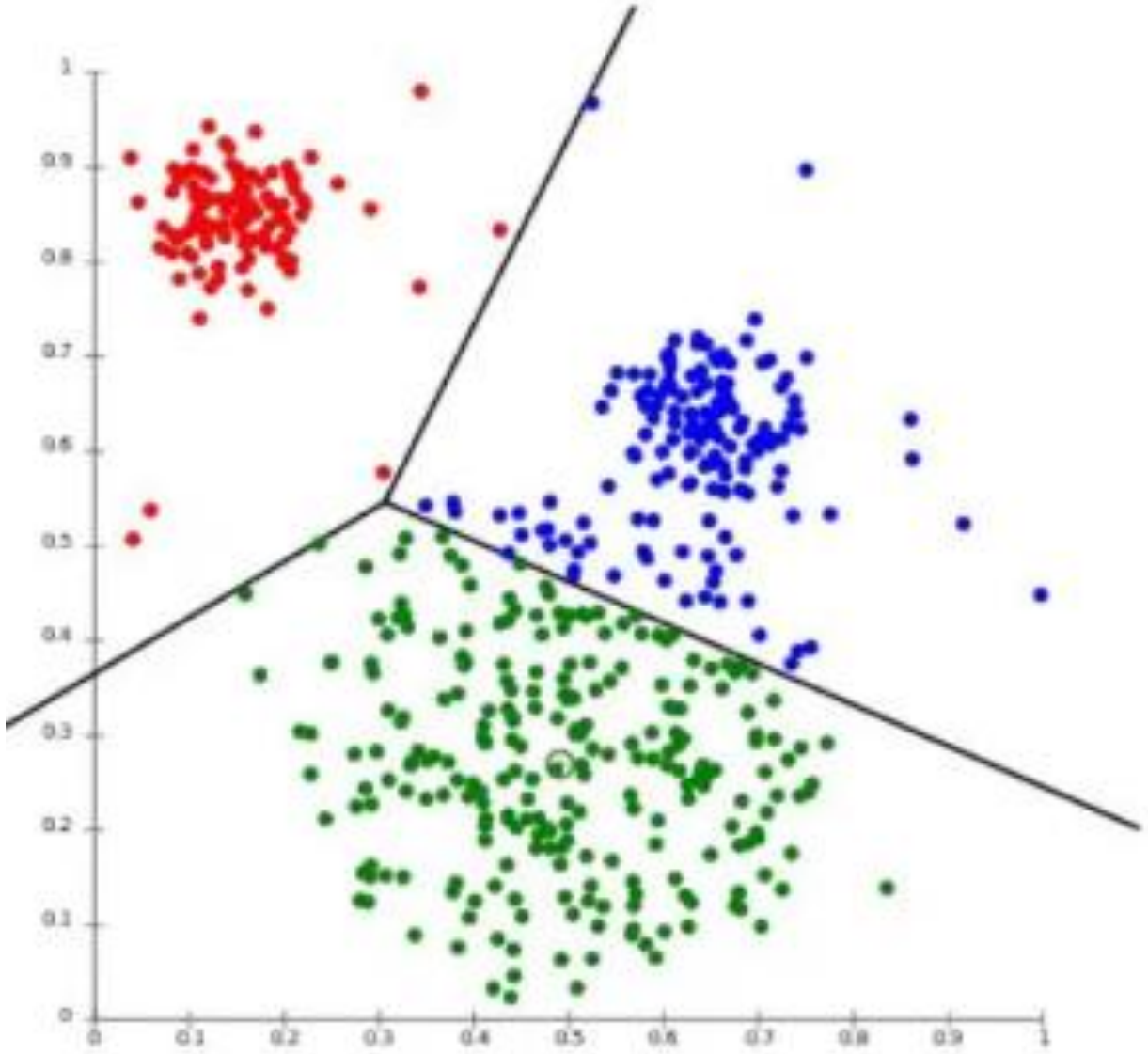
مرحله ۴ – سپس، مراحل زیر را تا زمانی که مرکز بهینه را پیدا کنیم، ادامه دهید. مرکز بهینه، انتساب نقاط داده به خوشه‌هایی است که دیگر تغییر نمی‌کنند.

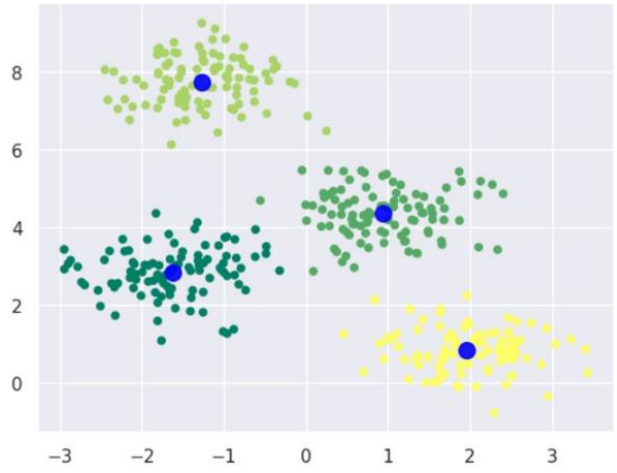
۴،۱ – ابتدا، مجموع مربعات فاصله بین نقاط داده و مراکز محاسبه می‌شود.

۴،۲ – اکنون، باید هر نقطه داده را به خوشه‌ای که از خوشه دیگر (مرکز خوشه) نزدیک‌تر است، اختصاص دهیم.

۴،۳ – در نهایت، مراکز خوشه‌ها را با گرفتن میانگین تمام نقاط داده آن خوشه محاسبه می‌کنیم.

K-Means Clustering



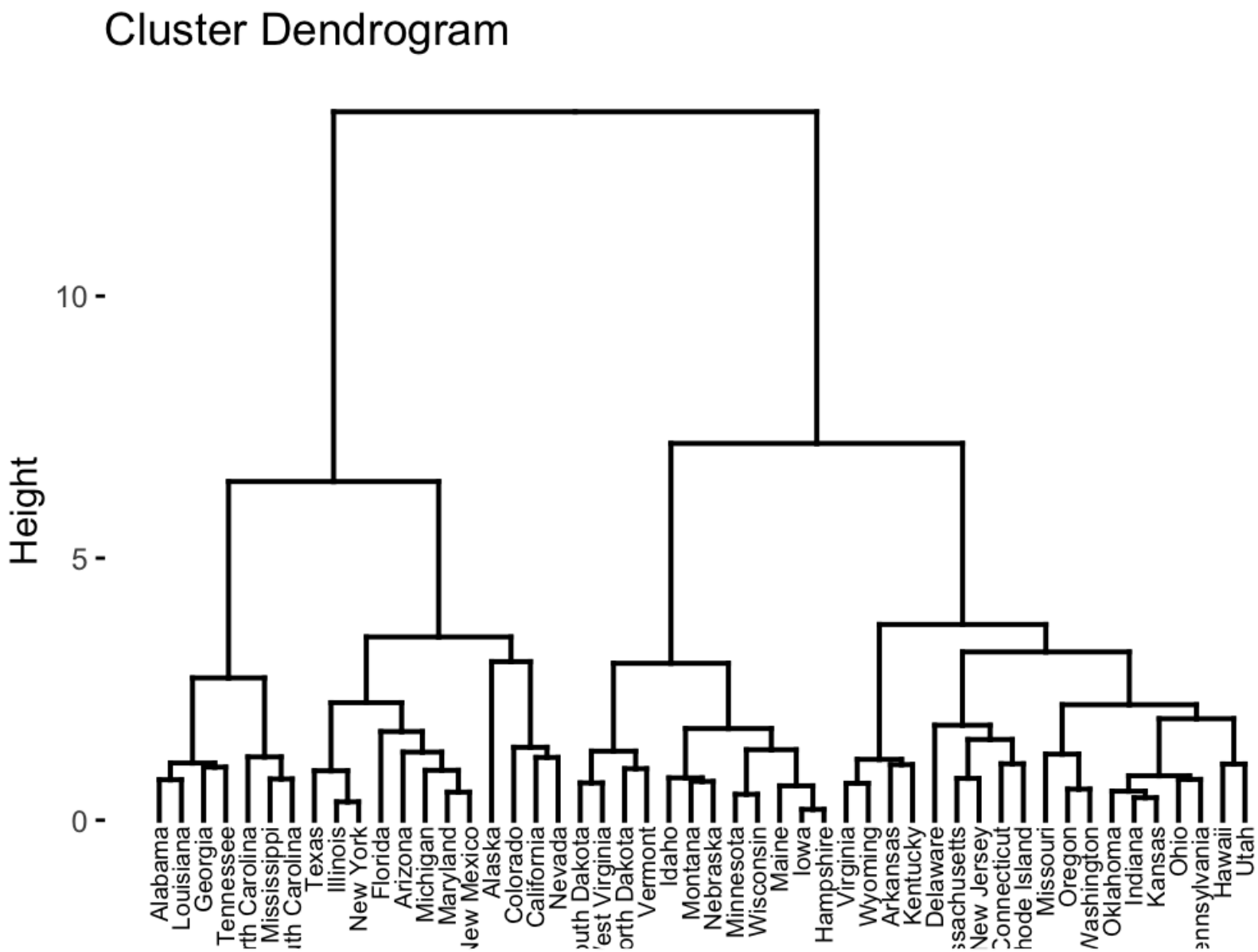


کاربردها:

- تقسیم‌بندی مشتریان
- بخش‌بندی تصویر
- تشخیص نقاط پرت و ناهنجاری

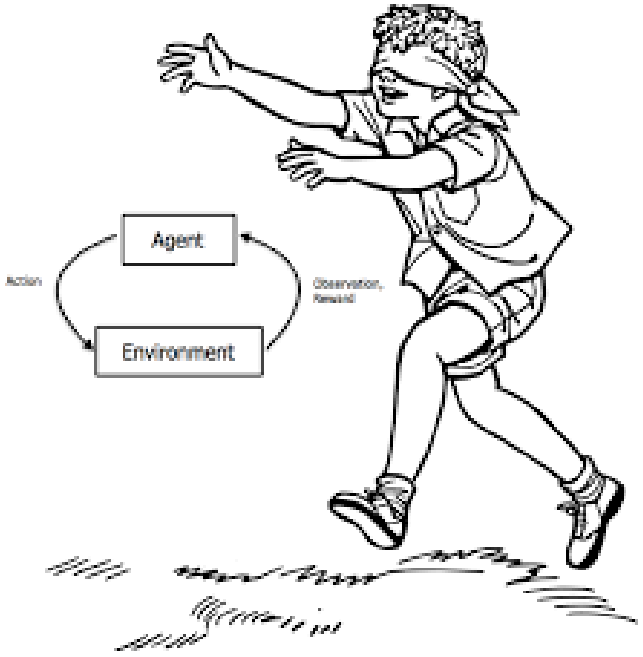
❑ Fuzzy c-means clustering

خوشه بندی سلسله مراتبی



یادگیری تقویتی

- یادگیری یک سیاست است، یعنی دنباله ای از اقدامات صحیح (با اخذ پاداش) برای رسیدن به هدف

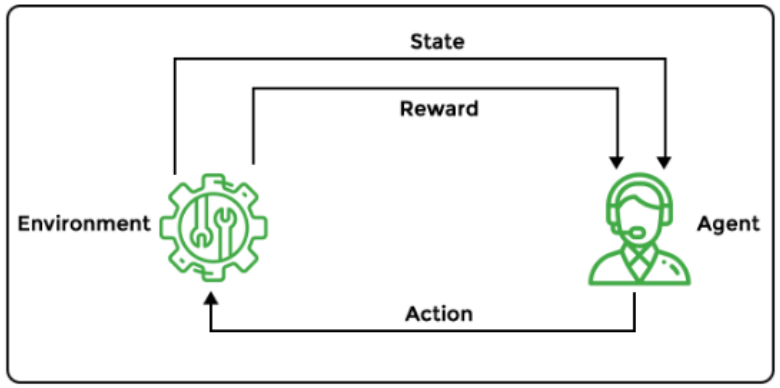




یادگیری تقویتی

عناصر سیستم یادگیری تقویتی

- عامل
- محیط
- سیاست
- سیگنال پاداش
- تابع ارزش
- مدل



Markov Decision Processes(MDP)

یک چارچوب ریاضی برای مدل‌سازی تصمیم‌گیری در محیطی با حالت‌ها، اقدامات، پاداش‌ها و احتمالات

یادگیری تقویتی

مزایای یادگیری تقویتی

- یادگیری تقویتی نیازی به دستورالعمل‌های از پیش تعریف شده و مداخله انسانی ندارد.
- مدل یادگیری تقویتی می‌تواند با طیف وسیعی از محیط‌ها از جمله ایستا و پویا سازگار شود.
- یادگیری تقویتی می‌تواند برای حل مسائل تصمیم‌گیری، پیش‌بینی و بهینه‌سازی، مورد استفاده قرار گیرد.
- مدل یادگیری تقویتی با کسب تجربه و تنظیم دقیق، بهتر می‌شود.

معایب یادگیری تقویتی

- برخی از معایب یادگیری تقویتی عبارتند از:
- یادگیری تقویتی به کیفیت تابع پاداش بستگی دارد، اگر طراحی ضعیفی داشته باشد، مدل نمی‌تواند عملکرد بهتری داشته باشد.
- طراحی و تنظیم یادگیری تقویتی می‌تواند پیچیده باشد و نیاز به تخصص دارد.

Any question?