

یادگیری ماشین با پایتون

فرزاد حسینیان

دانشگاه محقق اردبیلی

دانشکده علوم کامپیوتر

آبان 98

جلسه صفرم

مقدمات

دوره یادگیری ماشین با پایتون

رویکرد:

مفاهیم
پیاده سازی
تمارین

تمرکز بر روی پیاده سازی از پایه با زبان پایتون خواهد بود

منابع

دانشگاه استنفورد

:Andrew Ng

coursera.org/learn/machine-learning



دانشگاه تبریز
سیدناصر رضوی:



youtube.com/channel/UCNJavFPJ16jeMP3zeviAJbQ

پیش نیازها:

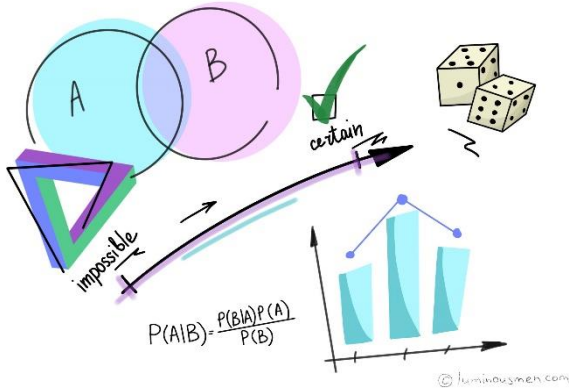
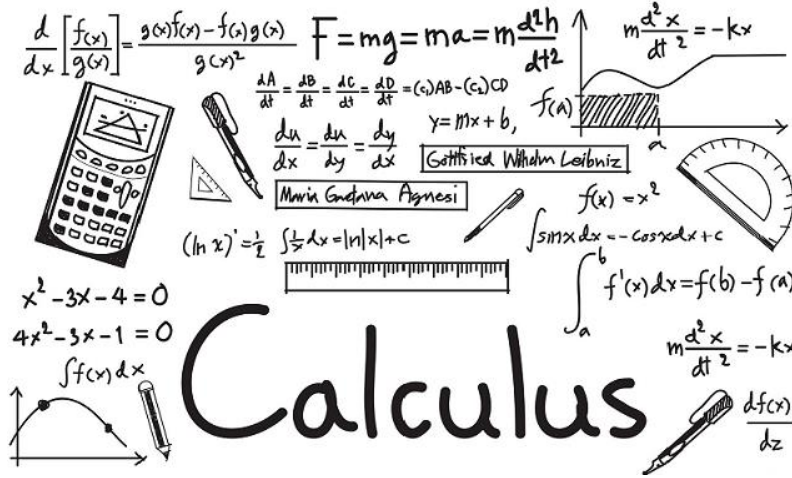
جبر خطی

حسابان

آمار

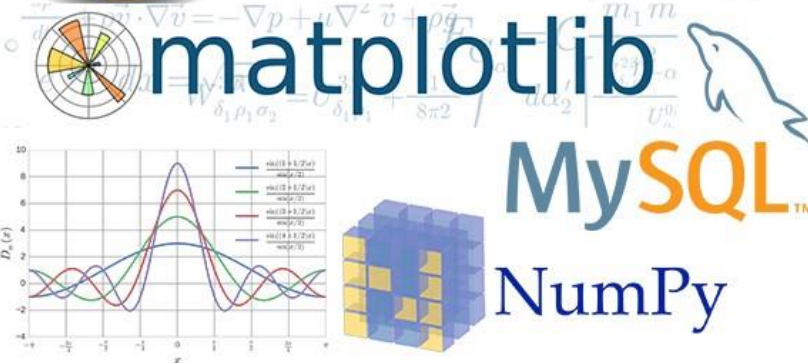
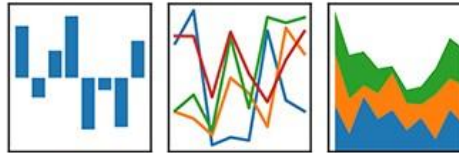
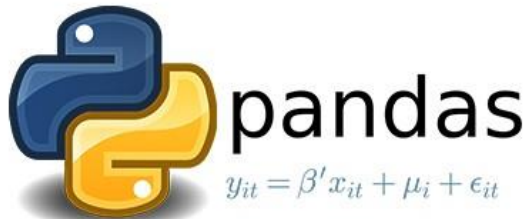
احتمال

پایتون



چرا پایتون؟

- سادگی
- پکیج های متعدد
- جامعه آماری بالا
- * در صورت استفاده از پایتون تمرکز اصلی در مطالعه به بررسی مفاهیم پایه یادگیری ماشین منتقل میشود.



تعریف یادگیری ماشین:

- برنامه نویسی برای بهینه سازی یک عملکرد با استفاده از داده ها و تجربیات گذشته
- دادن توانایی یادگیری به کامپیوتر بدون نیاز به برنامه نویسی صریح
- یادگیری ماشین در پی راهی برای ایجاد برنامه ای است که عملکرد را به صورت خودکار باتوجه به تجربیات قبلی ارتقا دهد

منظور از یادگیری:

تغییر نسبتاً پایدار در احساس، تفکر و رفتار که بر اثر **تجربه** ایجاد شده باشد

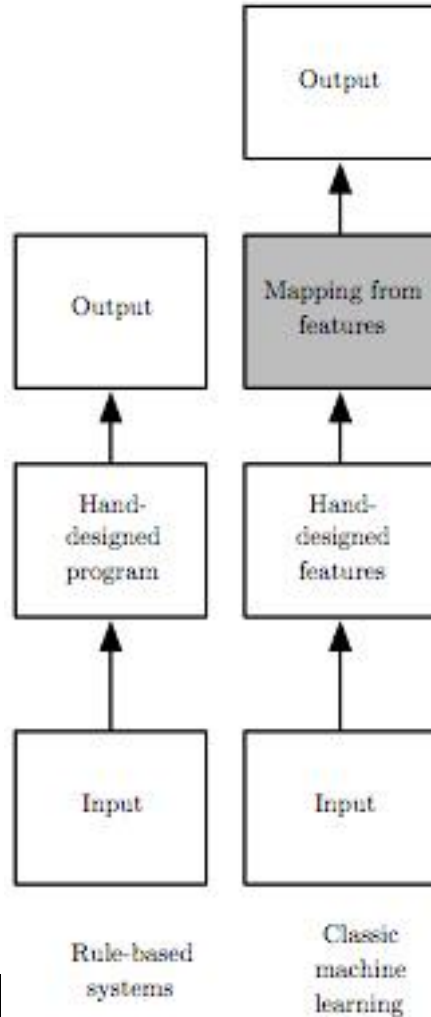
پیدا کردن مدلی که بتواند رفتار داده های مورد نظر را توجیح کند

:Tohm Mitchell(1997)

“A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at task in **T**, as measured by **P**, improves with experience **E**.”

"یادگیری ماشین به دنبال راهی برای برای ایجاد برنامه ای است که عملکرد را به صورت خودکار و با توجه به تجربیات قبلی ارتقا دهد."

برنامه نویسی استاتیک و یادگیری ماشین:



نگاشت:

در ریاضیات اگر به ازای هریک از اعضای یک مجموعه، تنها یک عضو در مجموعه ای دیگر وجود داشته باشد می گوییم نگاشتی از مجموعه اول به مجموعه دوم وجود دارد. به عبارت دیگر به ازای هر عنصر x در یک مجموعه تنها یک عنصر به نام $f(x)$ در مجموعه ای دیگر وجود خواهد داشت.

برنامه نویسی data-driven یا مبتنی بر دیتا:

برنامه‌نویسی داده محور یک پارادیم برنامه نویسی است که در آن دستورهای برنامه ، داده‌ها را برای همسان‌سازی و پردازش‌های مورد نیاز به جای تعریف یک توالی مراحل توصیف می‌کنند

پارادایم: شیوه برنامه نویسی

یک زبان برنامه‌نویسی می‌تواند یک یا چند شیوه برنامه‌نویسی را پشتیبانی نماید. برای مثال، برنامه‌های نوشته شده با سی++ می‌توانند کاملاً به صورت رویه‌ای باشند یا کاملاً منطبق بر شیوه برنامه‌نویسی شیء‌گرا که در تضاد کامل با شیوه رویه‌ای است بوده یا حتی حاوی عناصری از هر دو شیوه باشند. تصمیم‌گیری برای چگونگی استفاده از عناصر شیوه‌های برنامه‌نویسی بر عهده طراح برنامه یا برنامه‌نویس می‌باشد.

چرا یادگیری ماشین؟

برای حل مسائل بر روی یک کامپیوتر نیاز به یک الگوریتم داریم. برای برخی مسائل نمیتوان یک الگوریتم نوشت. مانند تشخیص ایمیل اسپم از ایمیل غیر اسپم. اما نمونه های زیادی از داده در اختیار داریم.

در جاهایی که نمیتوانیم مستقیماً برنامه مورد نظر را بنویسیم، نیاز به یادگیری داریم که با یکسری داده آموزشی یا تجربه صورت میپذیرد.

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

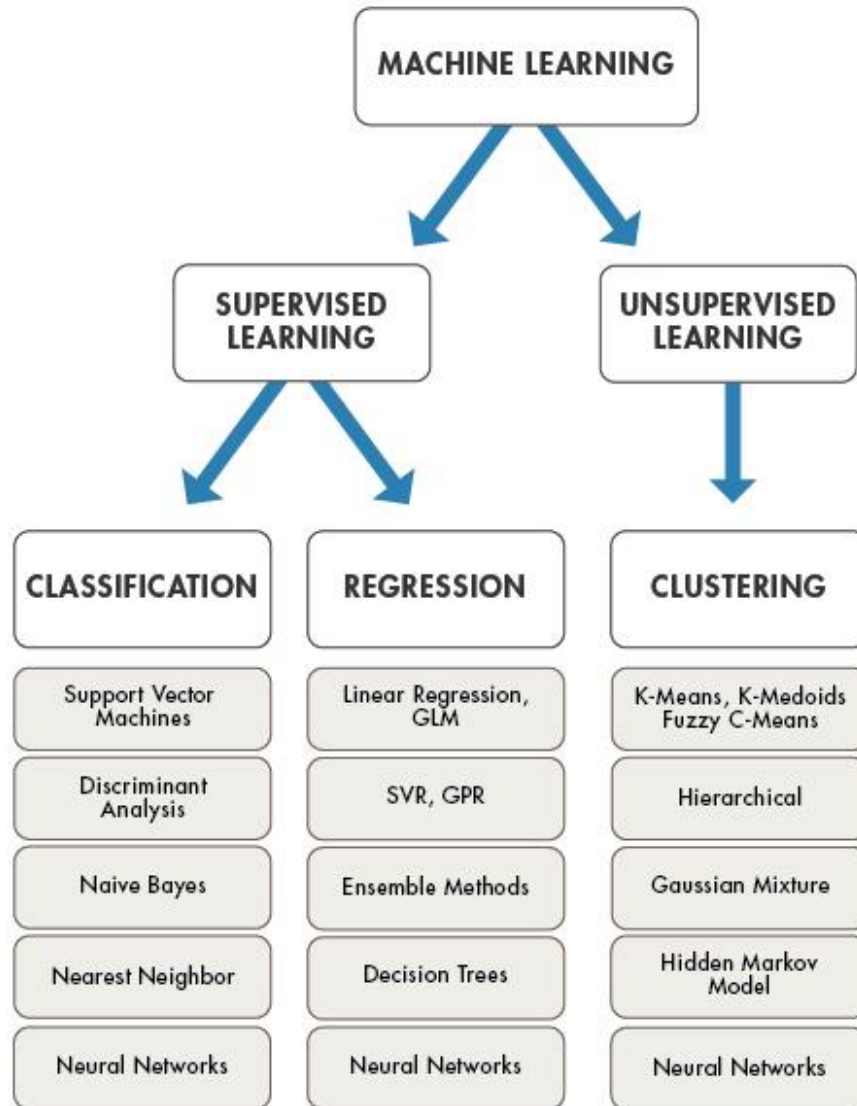
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



مزایای استفاده از یادگیری ماشین:

- هزینه پردازی بر مرحله آموزش است
- قابلیت تعمیم دارد
-

انواع روش های یادگیری ماشین:

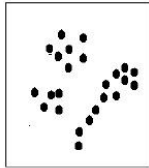


یادگیری بانظارت: برای هر داده x یک جواب (برچسب) y داریم
 قیمت خانه با استفاده از ویژگی های خانه، بدخیم یا خوش بودن سرطان و...

یادگیری بدون نظارت: برای داده ها برچسبی نداریم
 تحلیل داده های شبکه های اجتماعی، دسته بندی خبرهای وب سایت ها، تحلیل
 داده های ستارگان و...

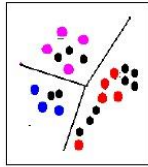
Types of Grouping

**Unsupervised
Learning**



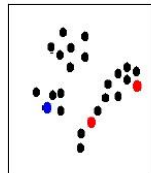
We know nothing
about data structure

**Supervised
Learning**



We know well
data structure

**Semi-Supervised
Learning**



We know something
about data structure

یادگیری نیمه نظارت شده:

پیشنهاد جمله یا ایموجی در چت

یادگیری تقویتی

خودرو های خودران

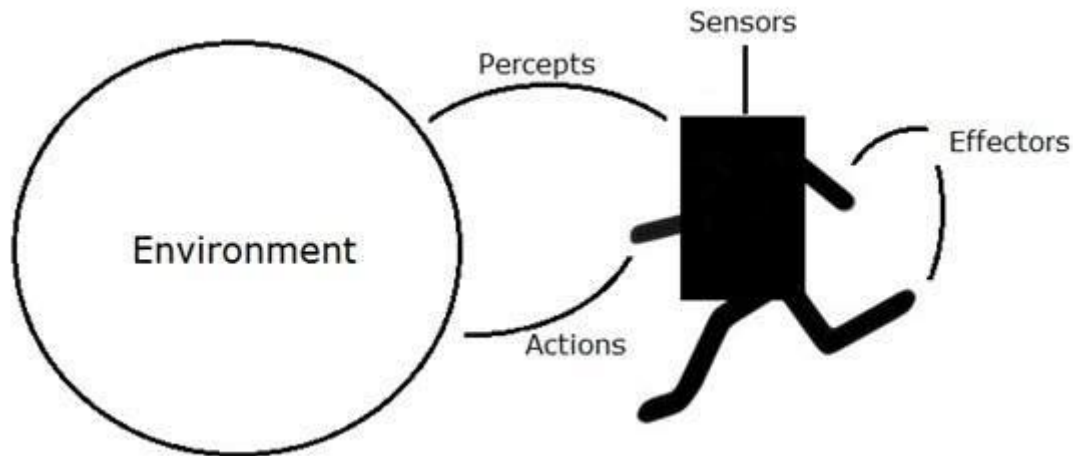
TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE
SOME REALLY GOOD POINTS.

I'M ... NOT EVEN SURE
WHO I AM ANYMORE.

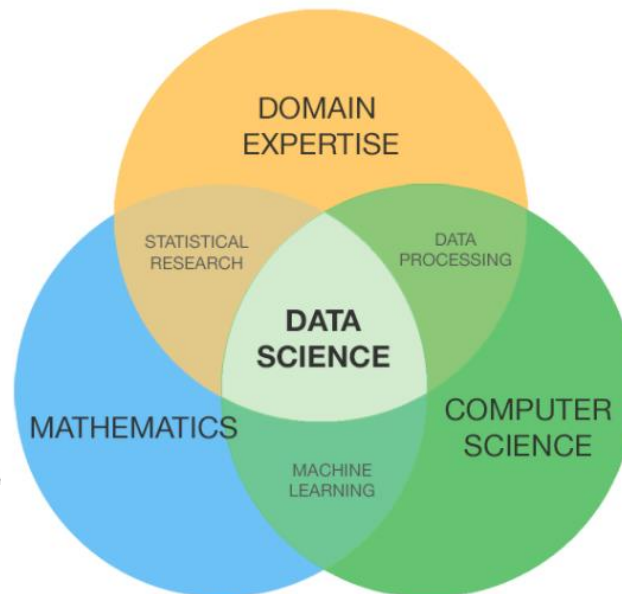
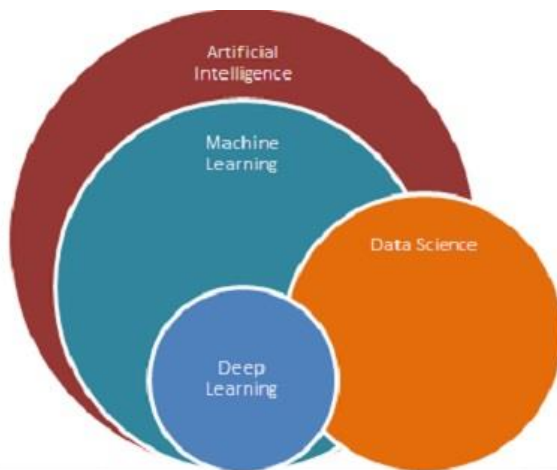


هوش مصنوعی:



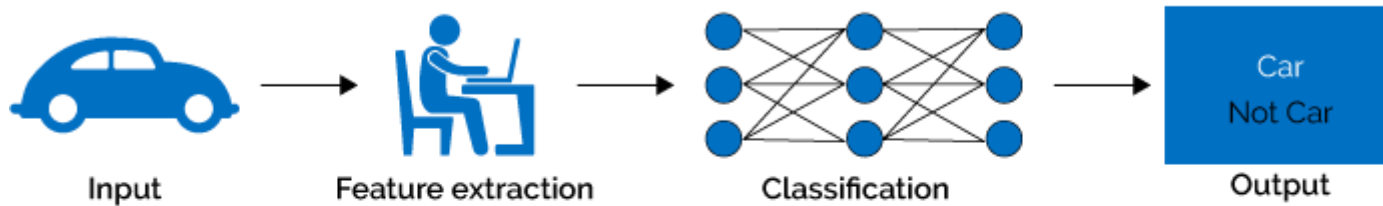
Data Science & Data mining

“The process of **discovering patterns** in data. The process must be **automatic** or (more usually) semiautomatic. The patterns discovered must be **meaningful** in that they lead to some advantage—e.g., an economic advantage. The data is invariably present in substantial quantities.”

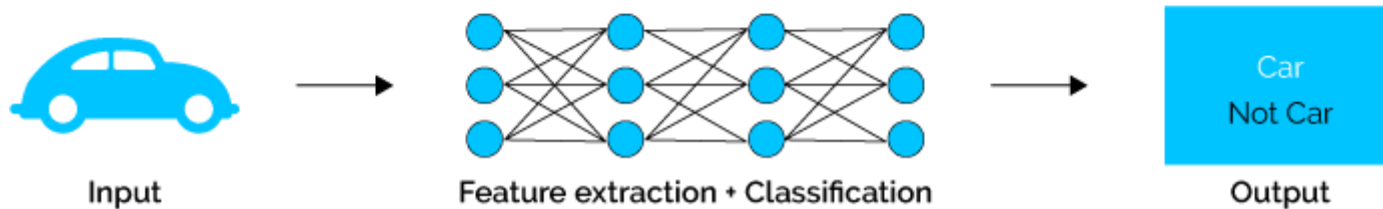


یادگیری عمیق DEEP learning

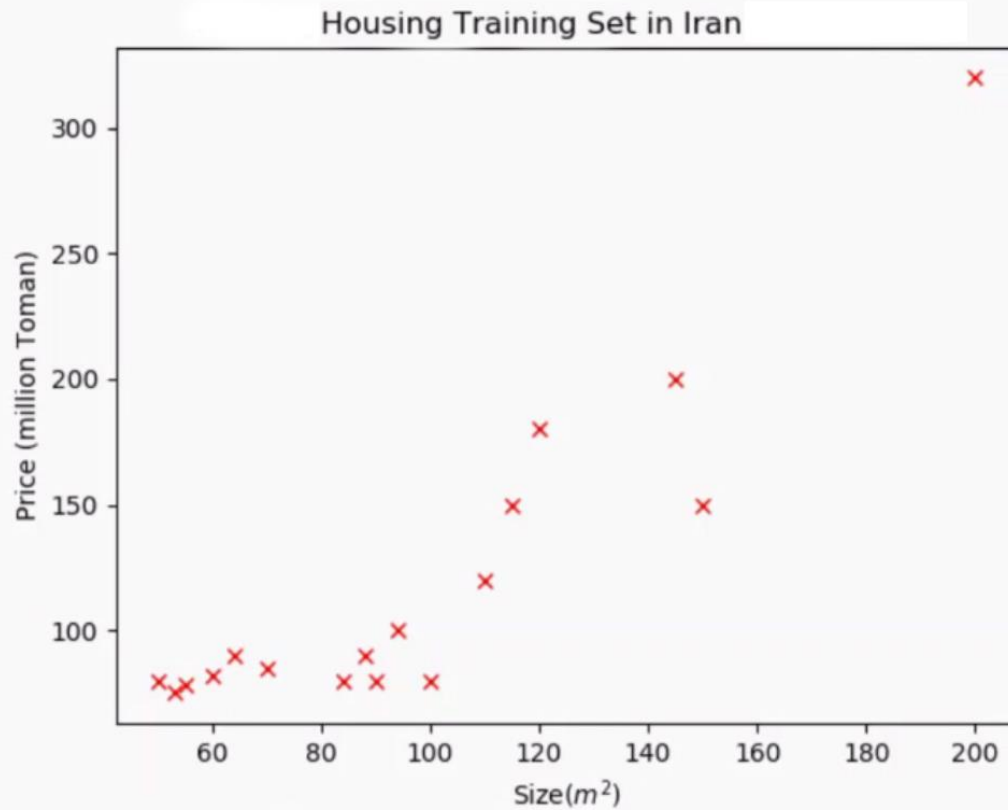
Machine Learning



Deep Learning



رگرسیون خطی:



Training set of housing prices (Portland, OR)

Size in feet ² (x)	Price (\$) in 1000's (y)
→ 2104	460
1416	232
→ 1534	315
852	178
...	...

} m = 47

Notation:

- m = Number of training examples
- x 's = "input" variable / features
- y 's = "output" variable / "target" variable

(x, y) - one training example

$(x^{(i)}, y^{(i)})$ - i th training example

$$\begin{cases} x^{(1)} = 2104 \\ x^{(2)} = 1416 \\ y^{(1)} = 460 \end{cases}$$

Andrew Ng

Hypothesis: مدلی که الگوریتم یادگیریمان براساس
دیتاها یاد میگیرد.

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Parameters:

$$\theta_0, \theta_1$$