

MACHINE LEARNING



دوره جامع یادگیری ماشین

قسمت پنجم، یادگیررهار ماشین: براساسرطوع نظارت

انواع سیستم های یادگیری ماشین ۱




DataTalk.ir

Created by : Ali Arabshahi

Contact us : [Linkedin.com/in/mrAliArabshahi](https://www.linkedin.com/in/mrAliArabshahi)

انواع سیستم های یادگیری ماشین

دسته بندی های متعددی برای سیستم های یادگیری ماشین بر اساس شاخص های مختلف وجود دارد که ما در چند جلسه پیش رو به سه تا از معروف ترین اون ها می پردازیم.

بر اساس با و یا بدون ناظر بودن یادگیری ماشینمون (supervised - unsupervised) 
بر اساس این که آیا یادگیریمون به طور پیوسته اتفاق می افتد (Online) یا با فواصل زمانی و حجم داده های از پیش تعیین شده (batch) 
و در آخر بر اساس این که آیا یادگیریمون صرفا با مقایسه داده های جدید و داده های قبلی اتفاق می افتد (instance based) یا با ساخت مدل و شناسایی الگوی مشخص در داده ها (model based) 

البته این دسته بندی ها مطلق نیستند و یه ماشین هوشمند می تونه به صورت ترکیبی هم مجموعه ای از این ویژگی ها رو با هم داشته باشه. به عنوان مثال ماشین تشخیص ایمیل های اسپممون می تونه با ناظر باشه و یادگیریش به صورت آنلاین اتفاق بیفته و خروجی ها توسط مدلی از پیش تعیین شده ، پیشبینی بشه. (ترکیبی از هر سه دسته بندی بالا)

بخشید اگر یکم داره مباحث پیچیده می شه اما بهتون قول میدم یکم که جلوتر بریم خیلی بیشتر این دسته بندی ها براتون روشن بشه . تو این فصل فقط می خوایم به نیم نگاهی به همه این دسته بندی ها بندازیم. اصلا نیازی نیست تو این مسائل خیلی عمیق بشین .

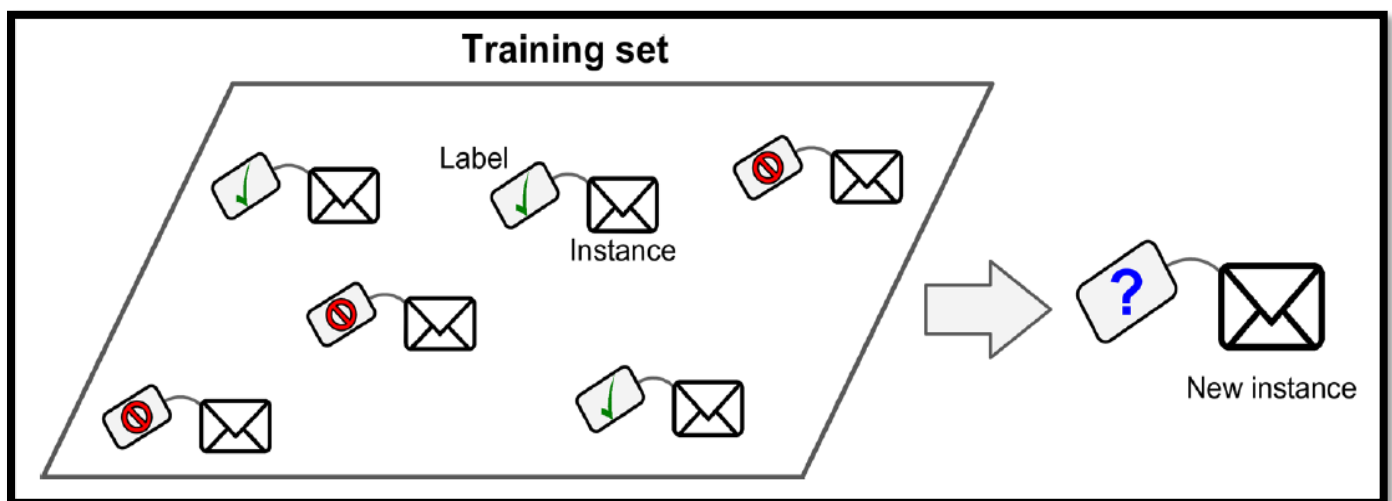
یادگیری بر اساس سطوح نظارت

بر اساس نوع نظارتی که در داده ها اتفاق می افتد، سیستم های یادگیری می توانند در چهار طبقه مختلف قرار بگیرند.

- 1) Supervised learning
- 2) unsupervised learning
- 3) semisupervised learning
- 4) and Reinforcement Learning

Supervised Learning

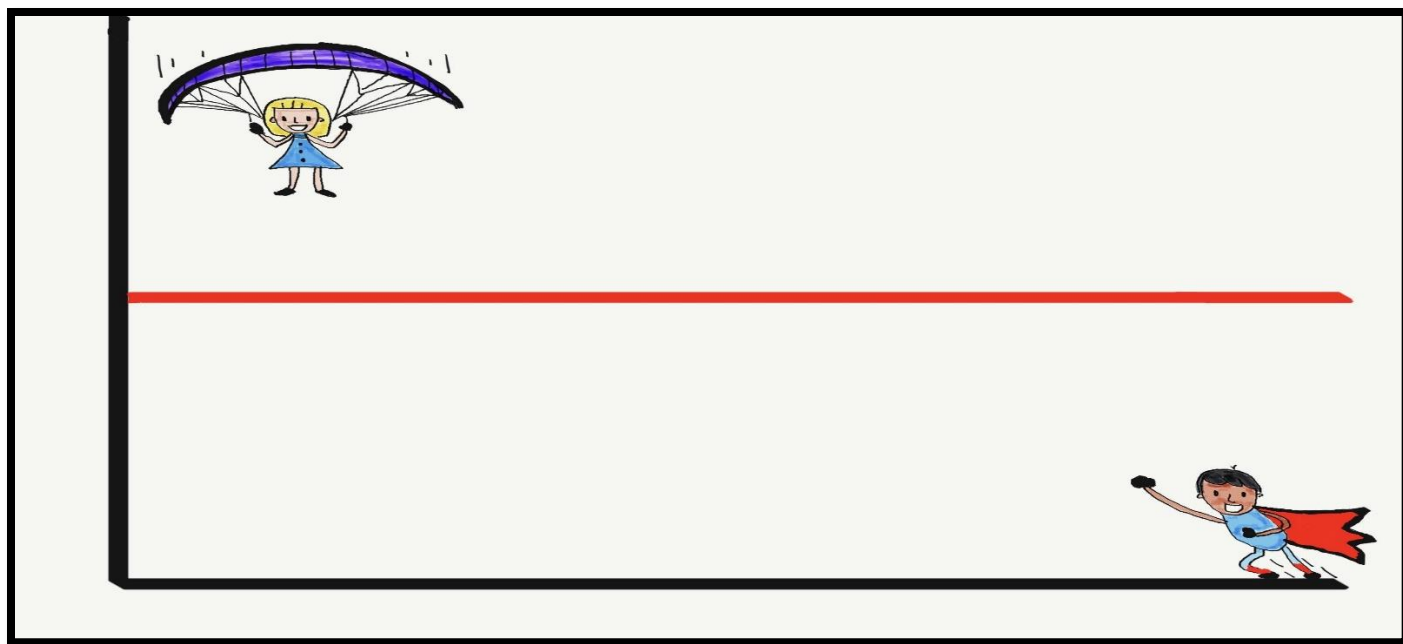
در یادگیری بانظر داده های آموزش (training set)، شامل لیبل هستن. یعنی به به مدل فهمونده می شن! خروجی مون مشخصه. مثلا یک مدل رگرسیون برای پیشبینی قیمت خانه رو در نظر بگیرین. میگین آقای مدلی که می خوای train بشی (یاد بگیره)، این خونه ها با این ویژگی ها (تعداد اتاق، متراژ و ...)، دارای **خروجی مشخص (فلان قیمت)** هستند. حالا بیا و یاد بگیر و قیمت خونه های جدید رو به من بگو (داده هامون لیبل دارن)



دیدین؟ 🤖 داده های مون هم یک سری X (ویژگی یا فیچر ها) دارن و هم Y (خروجی) .

یه نوع معروف دیگه ای از این دسته مسایل ، مسایل مربوط به دسته بندی (classification) هست. همون ماشین ایمیل اسپم می تونه مثال خوبی برای این مدل ها باشه. ما یک سری ایمیل داریم با ویژگی های مربوط به خودشون و می دونیم کدومشون اسپم و کدومشون ایمیل های عادی هستند (داده هامون لیبل دارن). **خروجی مشخصه!**

راستی یه نکته ی نه چندان خیلی با مزه راجع به **رگرسیون!** یه آقای به نام فرانسیس گالتون! داشت درباره ارتباط بین قد فرزندا با والینشون تحقیق می کرد. به طرز خیلی عجیبی مشاهده کرد که قد بچه های پدر و مادر های بلند قد! از اونا کوتاه تر میشه!!! بعد از کلی فکر کردن فهمید که قد بچه ها یه جور کشش خاصی به سمت میانگین کل قد ها داره. اسم این پدیده رو **regression to mean** یا تمایل و تعصب به میانگین نامید. و اینجوری شد که رگرسیون شد 😊 !

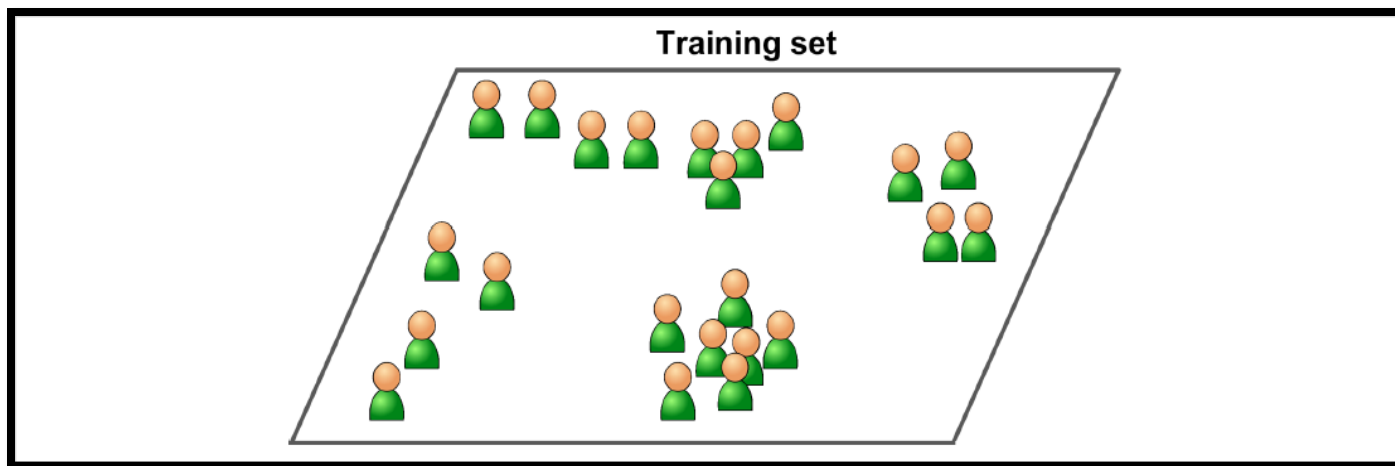


در پایین یک سری از مهم ترین الگوریتم های یادگیری ماشین بانظر رو مشاهده می کنید که در آینده قراره با همشون آشنا بشیم.

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks (البته بعضی هاشون)
- Unsupervised Learning

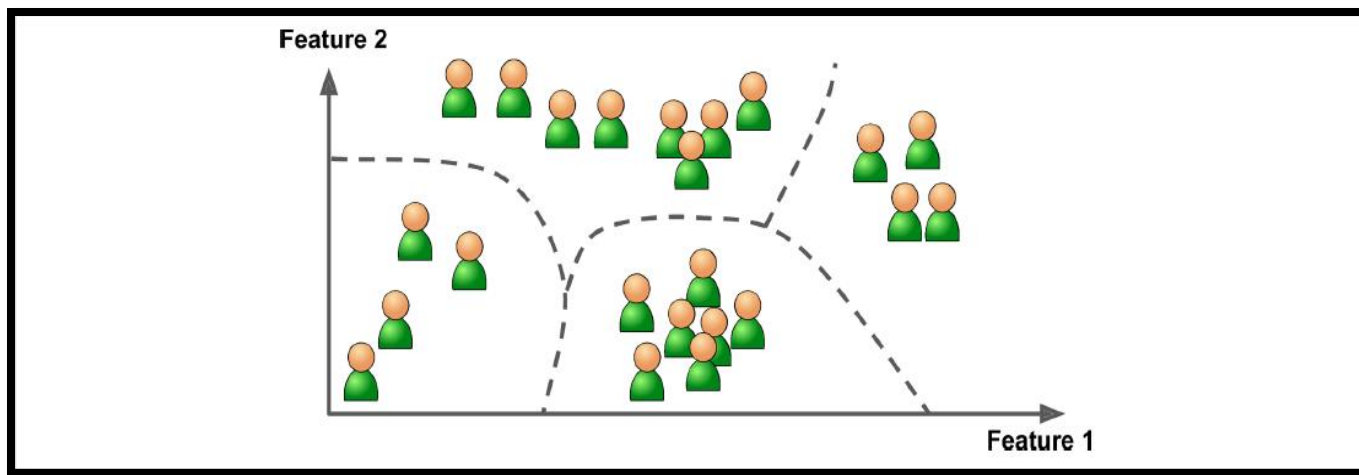


در یادگیری بدون ناظر همون طور که احتمالا خودتون هم می تونین حدس بزنین، داده هامون لیبل (Y) ندارن یا به تعبیر دیگه ماشینمون باید بدون معلم (ناظر) خودش یه چیزی یاد بگیره .



برای مثال فرض کنیم یه وبسایتی داریم و کلی دیتا از رفتاری که کاربرانمون از خودشون ثبت می کنن. شاید بخوایم طبقه بندی از کل کاربرانمون داشته باشیم که بهمون بگه ۴۰ درصد کاربران از مطالب خبری بیشتر خوششون می یاد و معمولا شب ها به سایتمون مراجعه می کنند و ۲۰ درصد هم دوست دارن به محتوای طنز وبسایتمون دسترسی پیدا کنن و معمولا صبح ها می یان و ۴۰ درصد دیگه هم انگار خودشونم نمی دونن به چی علاقه دارن و معمولا نصف شبها به سایت سر می زنن! (انگار بی خوابی میزنه به سرشون!) 😊

متوجه شدین؟ ما یک سری دیتا داریم و هیچ درکی ازشون نداریم. (در مثال قیمت خانه یا ایمیل اسپم ما می دونستیم هر داده چه معنی داره و به چه خروجی یی منجر میشه (قیمت، اسپم یا غیر اسپم و ...))

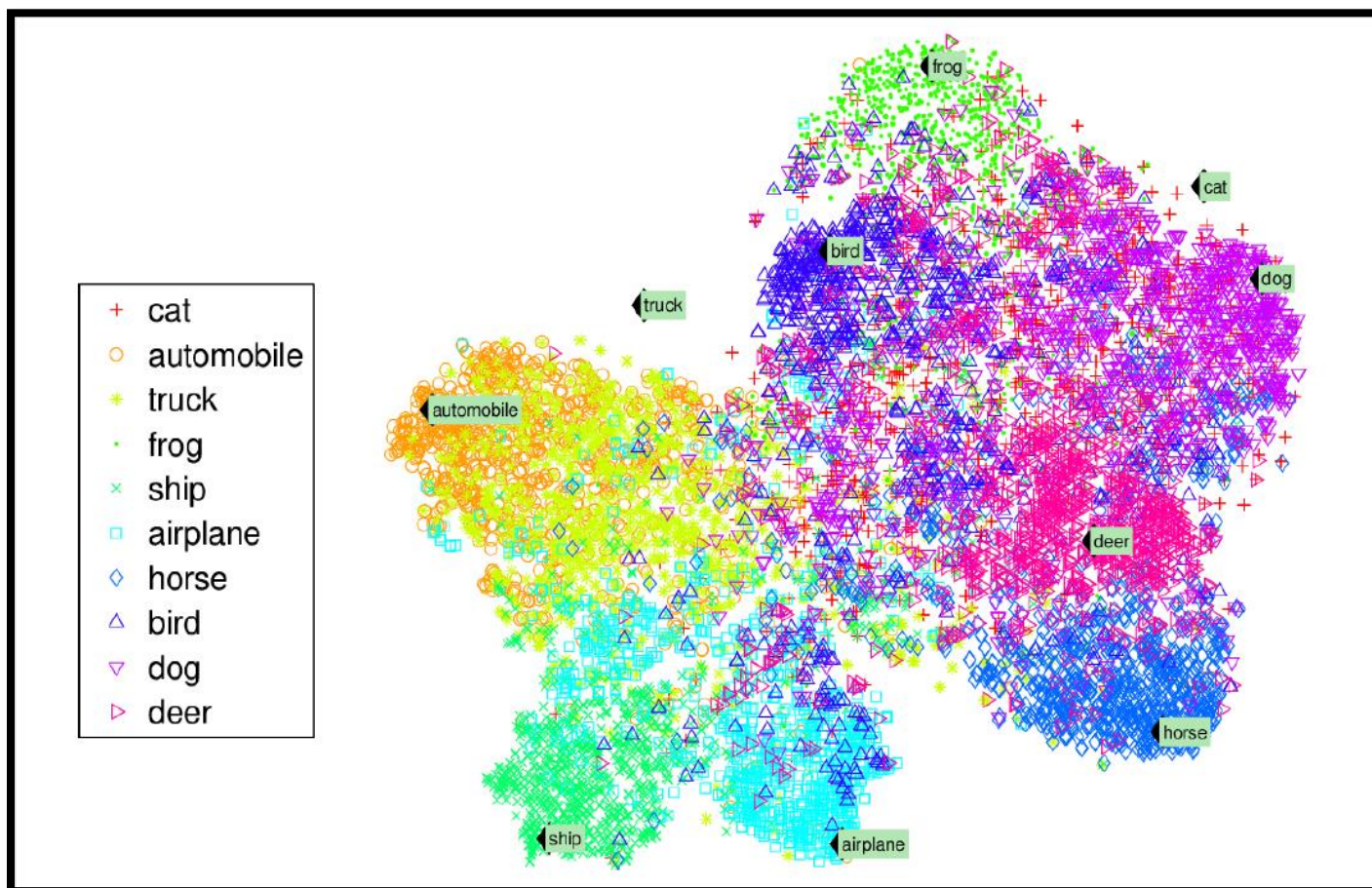


ماشین داده ها رو میگیره و یه دانشی بهمون میده! خیلی جالبه ها 😊

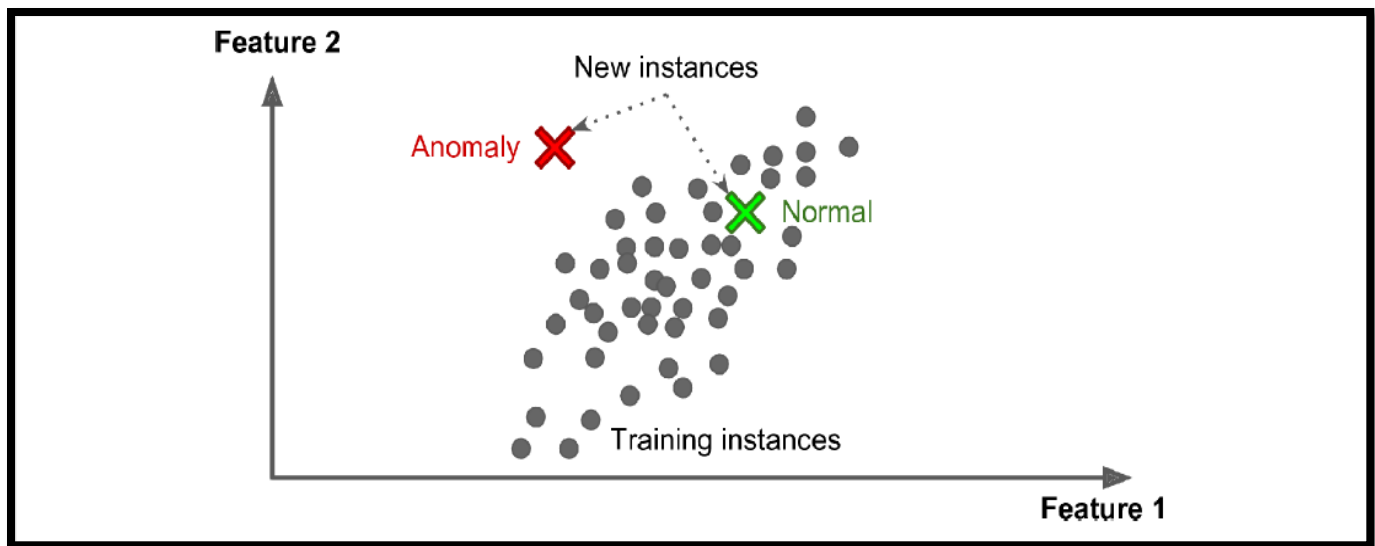
ما خودمونم نمی دونیم چی می خوایم اما ماشین کمکمون می کنه که یه چیز ارزشمندی به دست بیاریم. این است قدرت یادگیری ماشین! 🤖

مثال دیگه ای از الگوریتم های بدون ناظر، الگوریتم های کاهش ابعاد (dimensionality reduction) هستند. رویکردشون این شکلی یه که میان داده ها ی مرتبط به هم رو با هم ادغام می کنن. مثلا مقدار کیلومتری که ماشین ها کار کردن احتمالا با سن ماشین ها رابطه ی خیلی مستقیمی داشته باشه پس الگوریتم میاد و این دو تا فیچر رو تبدیل به یکی می کنه. این کار دو تا تا کاربرد خیلی خوب می تونه داشته باشن. یکی این که حجم دیتاهامون خیلی کم تر میشه (برای همین شاید این کار خوب باشه که حتی در زمانی که می خوایم از یه الگوریتم با ناظر استفاده کنیم، قبلش یه الگوریتم کاهش ابعاد رو روی داده هامون پیاده سازی کنیم). و حالت دیگه هم برای داستان مصور سازی داده ها استفاده میشه. ما آدم ها یه محدودیت کوچیک داریم و اون این که بیشتر از ۳ بعد رو نمیتونیم ببینیم! دیتاهایی داریم که از ابعاد خیلی زیادی تشکیل شدن (مثلا سن، وزن، قد و ...) و میخوایم همه ی

اون ها رو در قالب یک نمودار ببینیم. خب با اجرای این الگوریتم خیلی راحت می تونیم همه ابعادمون رو تبدیل به دو یا ۳ بعد کنیم و بعد به نمایش در بیاریم.



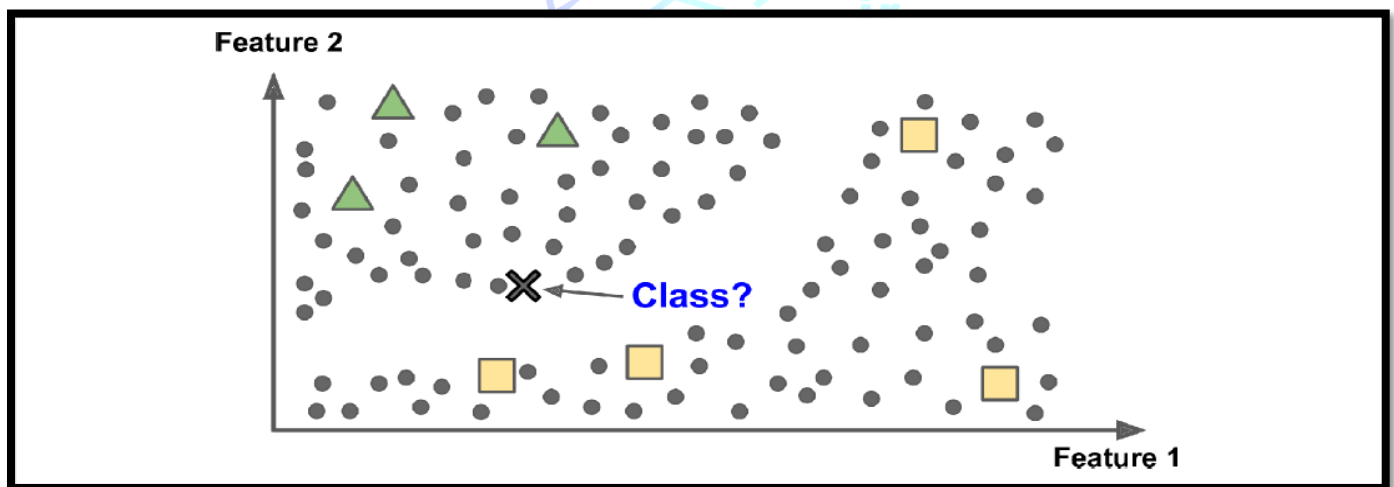
نوع دیگه ای از الگوریتم های بدون ناظر ، الگوریتم های تشخیص نا هنجاری (anomaly detection) هستند. اگر بخوایم از کل تراکنش های بانکی که داریم اونایی رو که نسبت به اکثر داده ها رفتار یا شکل متفاوت تری رو دارن (که احتمالا نشون دهنده ی اختلاس گرای خودمون هستن!) شناسایی کنیم یا مثلا در یک خط تولید محصولاتی رو که عادی نیستن رو بشناسیم ، می تونیم از این نوع الگوریتم ها استفاده کنیم.



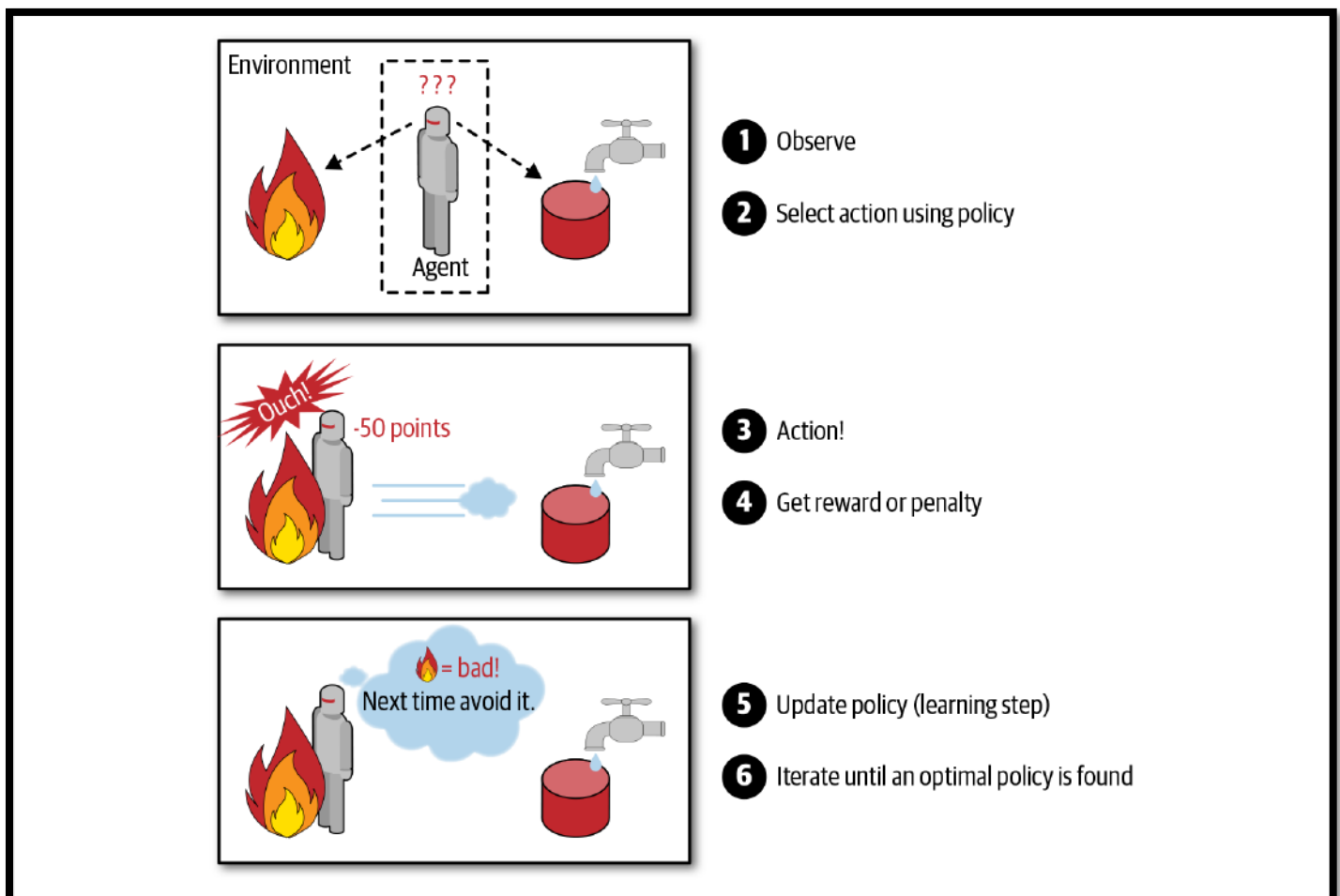
مشابه این الگوریتم ها، novelty detection یا تشخیص موارد جدید و نو ظهور هست. فرقی با anomaly detection اینه که در اونا داده های آموزش مون باید خیلی تمیز و یکدست باشن، مثلاً نوع خاصی از گُل، و در نهایت الگوریتم، هر گُل جدیدی حتی در صورتی که شبیه گُل هایی که به مدل آموزش دادیم باشن رو شناسایی می کنه.

و مثال دیگری که از این رویکرد (بدون ناظر) براتون می زنیم الگوریتم های یادگیری قانون وابستگی (association rule learning) هستند. هدفشون شناسایی یک سری همگرایی هایی از ویژگی های مختلف داده ها نسبت به هم هست. فکر کنید که مالک یک سوپرمارکت هستین. پیاده سازی این الگوریتم بهتون می فهمونه که اطلاعات فروش حکایت از این داره که مشتری هایی که چیپس می خرین احتمالاً تمایل به خرید ماست موسیر هم دارن (نگین که ندارین! لااقل یک بار امتحانش کنید 😊). پس چیپس ها رو به جایی بزارین که به قفسه ماست نزدیک تر باشه !

یادگیری نیمه یا شبه نظارت شده بیشتر زمانی استفاده میشه که ما داده هایی داریم که بعضی هاشون با لیبل و بعضی هاشون بدون لیبل هستند. از اون جایی که برای داده هایی با حجم زیاد لیبل زدن شاید کار خیلی زمانبری باشه این الگوریتم ها کشف شدن! اینستاگرام از این روش استفاده می کنه؛ چجوری وقتی عکس خانوادگی رو آپلود می کنیم به صورت خودکار نرم افزار اسم اشخاص مختلف رو برامون میاره؟ در قدم اول اینستاگرام میاد و اون افراد رو در عکس های دیگه تشخیص می ده، بر اساس شباهت ها و ... و در واقع برای هر فرد طبقه بندی مخصوص به خودش رو ایجاد می کنه . در قدم بعدی بر اساس یکی دو مورد لیبلی(اسم) که به افراد اختصاص داده شده(احتمالا توسط دوستانشون یا...)، به هم کدوم از عکس های دیگه هم بر اساس اون طبقه بندی های صورت گرفته اسم اختصاص می ده. در واقع رویکرد این الگوریتم ها تشکیل شدن از هر دو رویکرد یادگیری با ناظر(قدم دوم) و بدون ناظر(قدم اول)



رویکرد یادگیری تقویت شده تقریباً متفاوت هست با رویکرد های قبلی. روش کار به این صورته که ماشین رو رها می کنن تو یه بازی! و ماشین آزاده که هر تصمیمی می خواد بگیره. روش های مختلفی رو امتحان می کنه. بعضی از اون ها به امتیاز و بعضی از اون ها به جریمه ختم می شه. در نهایت سیستم بر اساس نتایجی که از رفتار های متعدد اش تجربه کرده، به یه سیاست (Policy) می رسه و با استفاده از اون در موقعیت های مشابه رفتاری رو از خودش بروز میده که منجر به افزایش امتیاز و یا به عبارت دیگه، عدم جریمه بشه.



و این جوریه که در خیلی از بازی های کامپیوتری جدید ما جرئت نمی کنیم سطح سختی بازی رو روی (hard) بگذاریم! ماشینی که داریم باهاش بازی می کنیم ریش هاش رو حسابی در این راه سفید کرده و اون می دونه بهترین تصمیم چیه. ای کاش خیلی از ما ها از این رویکرد الگو بگیریم. بدونیم زندگی ترکیبی هست از موفقیت و شکست های پی در پی . مهم اینه که این ماییم که در آخر برنده می شیم .

و تا جلسه ی دیگه، فعلاً 😊۱

