

# MACHINE LEARNING



## دوره جامع یادگیری ماشین

قسمت ششم، یادگیرنده‌ها را ماشین بر اساس کتابخانه بودی یا نبودن

## انواع سیستم های یادگیری ماشین ۲

DataTalk.ir

Created by : Ali Arabshahi

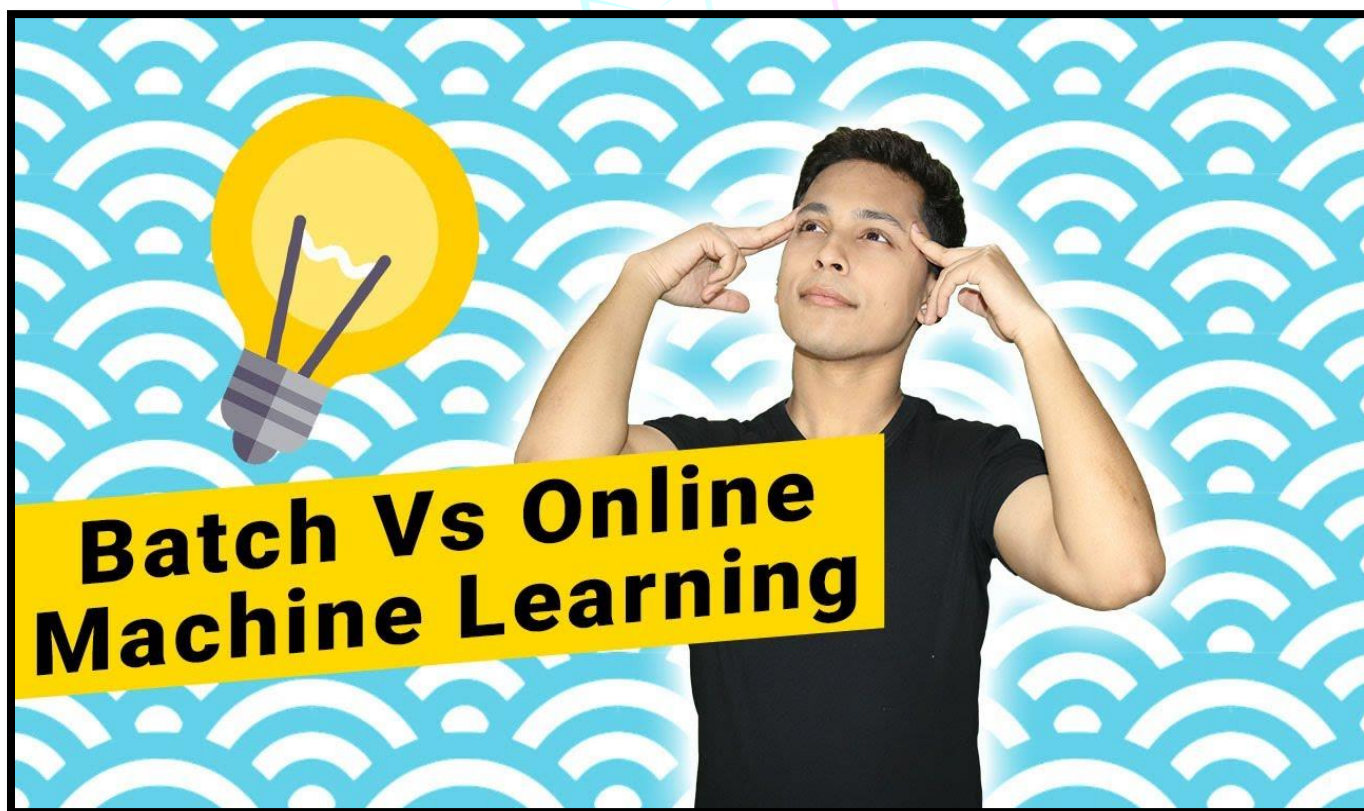
Contact us : [Linkedin.com/in/mrAliArabshahi](https://www.linkedin.com/in/mrAliArabshahi)

## انواع سیستم های یادگیری ماشین (ادامه)

در جلسه قبل به بررسی سیستم های یادگیری ماشین بر اساس سطوح مختلف نظارت پرداخته شد. در این قسمت مروری بر نوع دیگری از دسته بندی های رایج در این زمینه خواهیم داشت.

### Batch and Online Learning

یکی دیگر از معیار ها برای دسته بندی ماشین لرنینگ این است که آیا ماشین ما به طور پیوسته از دیتا هایی که از جریانی خاص وارد سیستم می شود، یاد می گیرد و یا این که یادگیری ما به طور گسسته و مقطعی اتفاق می افتد.



در سیستم های یادگیری دسته ای (Batch) ماشین قادر به یادگیری لحظه ای نیست. این یعنی چی؟ قبل از هر چیز ماشین ما آموزش داده میشه (Training). ممکنه این آموزش کلی طول بکشه و از منابع زیادی هم استفاده کنه. اما بعد از این همه داستان دیگه یه ماشین باهوش داریم که می تونه با دریافت داده های جدید بر اساس یادگیری قبلیش به ما خروجی مورد نظر رو تحویل بده. اما دقت کنین، بر اساس یادگیری قبلیش!

تو این دسته از سیستم ها، ماشین به صورت خودکار هوشمند نمی شه (مثلا فرض کنین اسپمر ها یک نوع جدیدی از ایمیل های اسپم رو ایجاد کردن؛ ماشین ما قادر به شناساییشون نیست!). البته می تونیم دوباره ماشینمون رو باهوش کنیم. چطوری؟

با متوقف کردن ماشین، فراهم کردن دوباره داده ها (هم داده های قدیمی و هم جدید) ، و آموزش دوباره ی ماشین. این کار اگرچه جواب میده اما ایراد های خاص خودش داره! مثلا ممکنه یادگیری ماشین مون کلی طول بکشه، یا نیاز به منابعی داشته باشه (لپ تاپ خفن دوستتون رو قرض گرفتن!) در عوض یک سری فوایدی هم داره دیگه، به عنوان نمونه زمانی که یادگیری تون تکمیل شد، ماشینتون بدون دردسر و با سرعت قابل قبول کار می کنه 😊

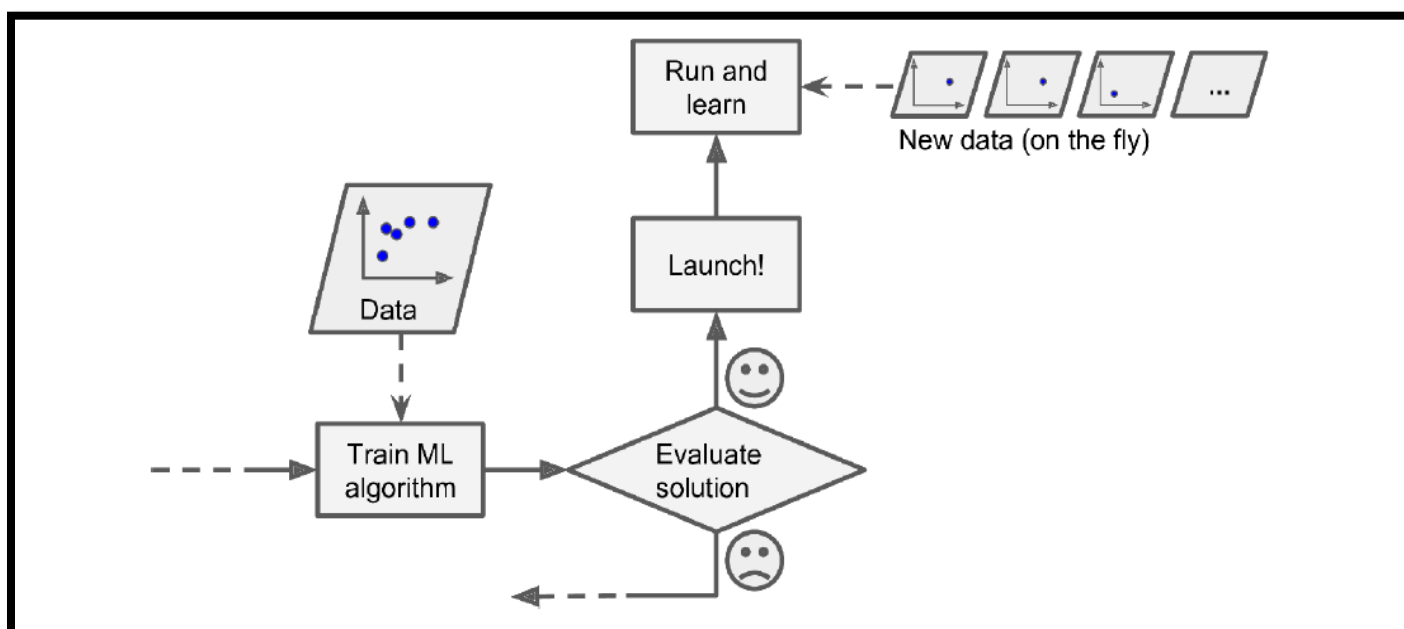
در ماشین لرنینگ چیزی به نام **خوب مطلق** و **یا بد مطلق** وجود نداره! وظیفه شما ی دیتا ساینس اینه که ببینین چه شرایطی بر مسئله حاکم هست و بر اساس نیازمندی هاتون از گزینه های مختلف که روی میز هست! استفاده کنین.

پس احتمالا تا الان متوجه شدین که زمانی که داده هامون به طور مرتب به روز می شن و تغییر می کنن (مثلا دیتا هایی که از بورس برای پیشبینی قیمت سهام میگیریم)، باید به سراغ نوع دیگه ای از سیستم های یادگیری ماشین بریم :

## Online Learning

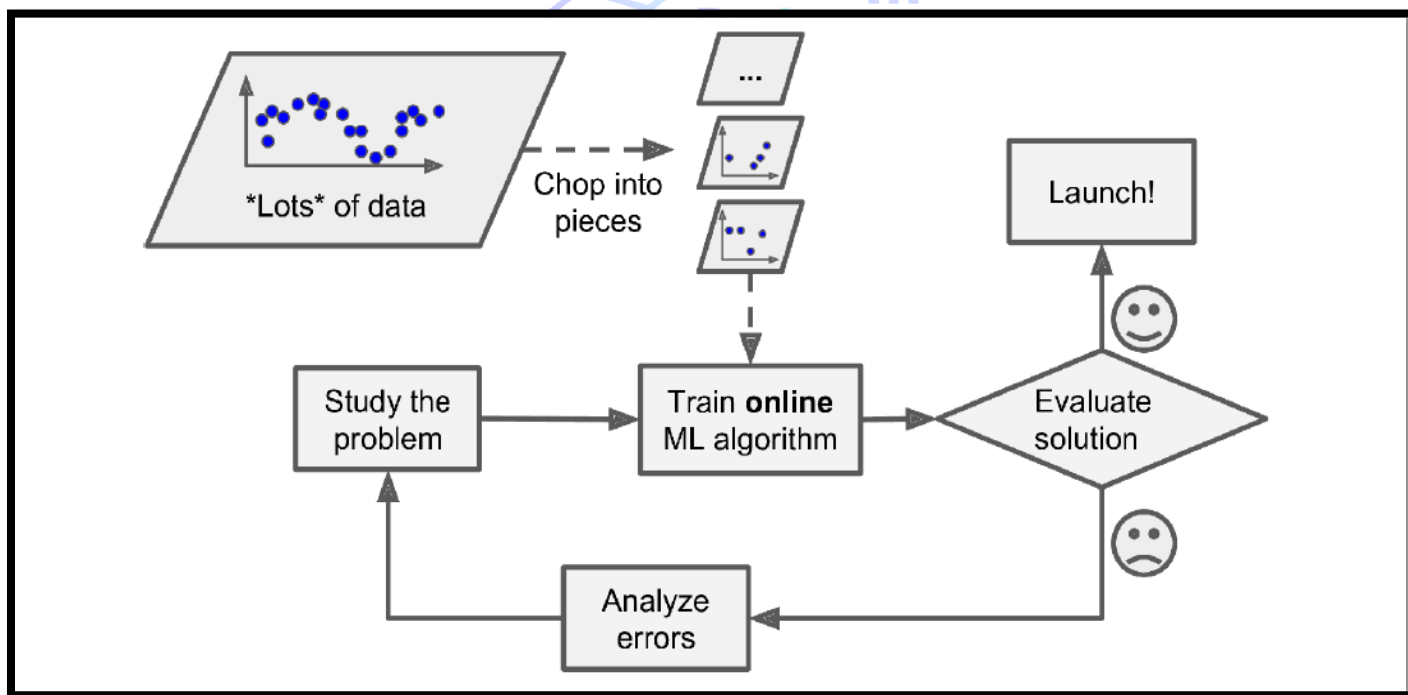
در این نوع از سیستم ها ماشین به طور لحظه ای یاد می گیره، چه این داده ها به صورت جداگانه وارد سیستم بشن و چه در قالب دسته های کوچک مثلا ۵ تا ۵ تا! (Mini Batches).

بنابراین رویکرد اصلی به این صورته که مدل (Train) میشه و اجرا می شه (کار میکنه) و در هر زمانی که داده های جدید وارد سیستم می شن یادگیریمون به روز میشه. ما آدمای باحال هم دقیقا همین سیستم رو داریم! فکر می کنیم کارمون درسته اما یه چیزای جدیدی تجربه می کنیم و تجدید نظر می کنیم! با هوش تر می شیم! روز به روز بهتر می شیم. اما اما از دست آدم هایی که سیستم یادگیری شون از نوع (Batch) هست و یادگیریشون رو به روز نمی کنن!

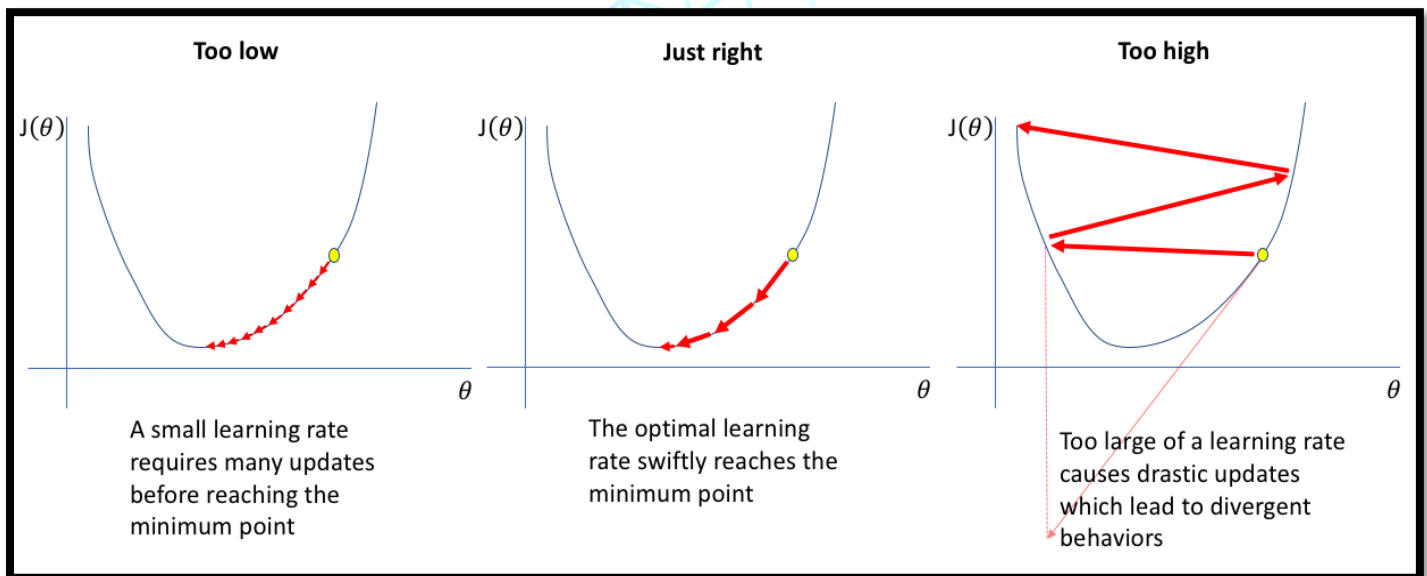


پس فهمیدیم که سیستم های یادگیری لحظه ای (Online) برای زمان هایی که داده هامون به طور مرتب به روز می شن (قیمت سهام) ، احتمالا گزینه ی بهتریه. خبر خوش این که می تونیم زمانی که داده های جدید اومدن و ماشین خودشو به روز کرد داده ها رو بندازیم دور! این باعث می شه کلی فضا نیاز نداشته باشیم (یادتون هست که در Batch Learning ، در هر باری که می خواستیم ماشین رو به روز کنیم نیاز بود کل داده همه اعم از قدیمی و جدید رو به مدل آموزش بدیم؟ پس همیشه باید دیتا هامون رو نگه داریم! 😊 )

یه نکته دیگه این که یادگیری لحظه ای (Online) برای مواقعی که حجم دیتامون خیلی زیاده و یک کامپیوتر به تنهایی نمیتونه از پس یادگیریش بر بیاد هم می تونه گزینه مناسبی باشه. (به این نوع از یادگیری ، out-of-core learning) میگن. ماشین یک قسمتی از دیتا رو یاد میگیره، بعد قسمت دیگه و همین طور تا زمانی که کل دیتا ها به مدل آموزش داده بشن.

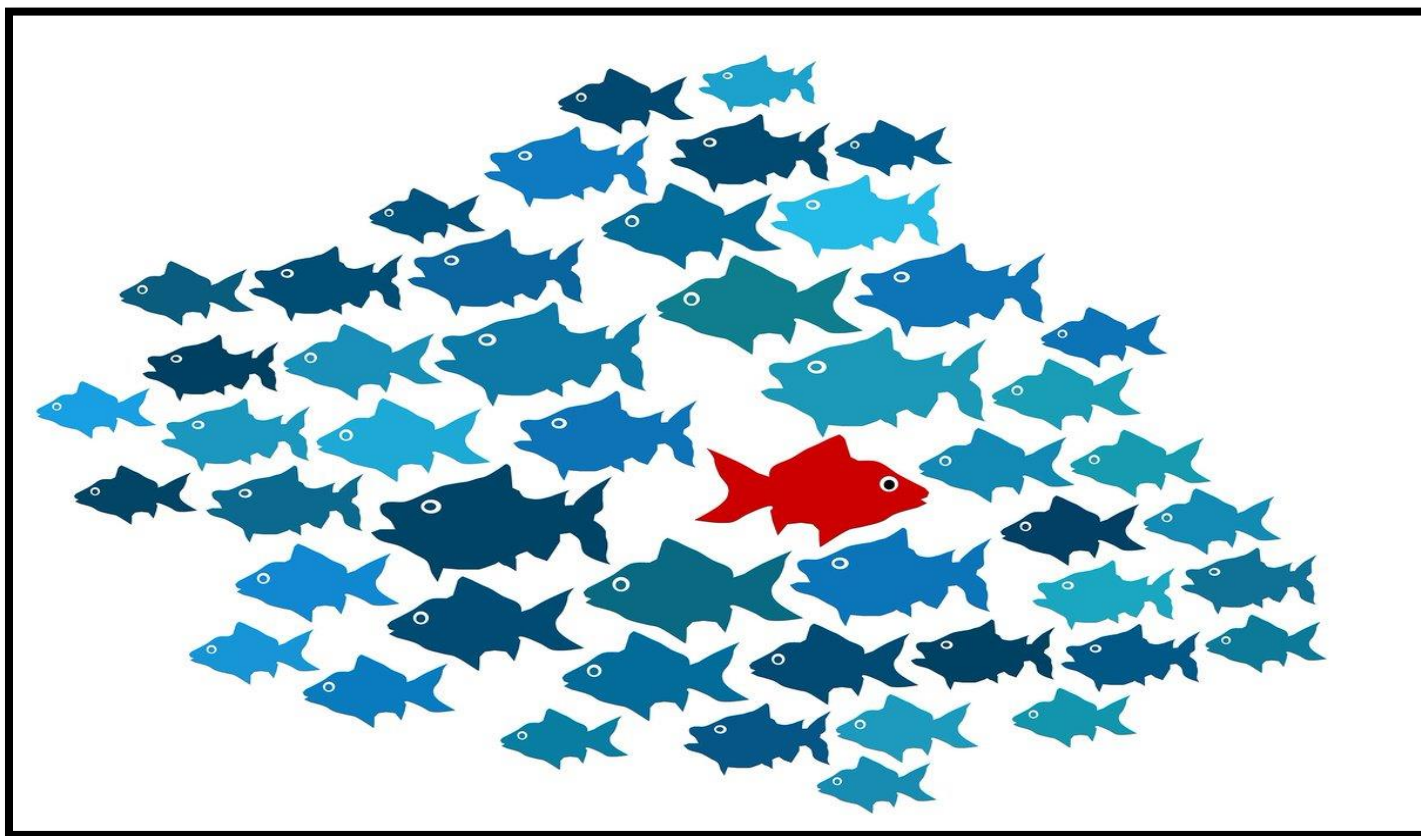


یک چالش خیلی مهم در سیستم های یادگیری آنلاین، انتخاب پارامتر یادگیری (Learning rate) هست. زیاد بودن این پارامتر به این معنی که ماشین ما به راحتی خودش رو با دیتا های جدید وقف می ده و از اونا تاثیر می پذیره اما از اون طرف تمایل داره که داده های قبلی رو هم با سرعت بیشتری فراموش کنه (بلانسبت شما، مثل این تازه به دوران رسیده ها که اصالت خودشون رو به راحتی با قرار گرفتن تو یه محیط جدید، فراموش می کنن!) و پایین بودن این پارامتر هم نشون از این داره که یادگیری مون با سرعت کمتری اتفاق می افته (ماشین دیر تر از داده های جدید تاثیر می پذیره) اما از اون طرف ماشینتون نسبت به داده های بدی که وارد سیستم می شن ، کمتر الگو می گیره. (ممکنه داده هامون بد آموزی داشته باشن! اما ماشین ما چشماشو به روی اونا می بنده! یا شاید بهتر باشه بگیم چشم هاش رو کامل به روی اونا باز نمی کنه 😊)





مهم ترین چالش در این نوع از سیستم ها اینه که اگر داده های غیر مناسب وارد سیستم بشن، مدل ممکنه از اونا یاد بگیره و این از کارایی ماشینمون می کاهه! راه حل پیشنهادی انتخاب مناسب پارامتر یادگیری و همچنین پایش مداوم داده های ورودی به مدل هست. می تونیم قبل از ورود داده ها به مدل اونا رو به ماشین دیگه ای بخورونیم! تا داده های نا متعارف اونا تشخیص داده و حذف بشن. آیا می تونین دیتای نامناسب در تصویر زیر رو پیدا کنین ؟ اونا رو پیدا و قبل از این که ماشینتون خل تر بشه، حذفش کنین. به همین راحتی!



ممونیم که تا اینجای کار با همراه بودین. به امید خدا در جلسه بعدی نوع آخر از دسته بندی های رایج در ماشین لرنینگ رو با هم بررسی می کنیم.

تا اون موقع، فعلا 😊