

# MACHINE LEARNING



## دوره جامع یادگیری ماشین

قسمت هفتم، یادگیرنده‌ها را ماشین بر اساس شباهت داده‌ها و یا مدل مصور بودن آن‌ها

## انواع سیستم‌های یادگیری ماشین ۳

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 این قسمت با آخرین نوع از دسته بندی های رایج یادگیری ماشین آشنا خواهیم شد.

### Instance-Based Versus Model-Based Learning

معیار دیگر برای دسته بندی سیستم های یادگیری ماشین بر این اساس است که چگونه تعمیم دادن (generalization) اتفاق می افتد؟ یعنی چی؟! 😊

حتما میدونین که یکی از بیشترین وظایف ماشین های هوشمند اینه که پیشبینی کنند. یعنی اول یک سری دیتا رو به ماشین آموزش بدیم (training data) و بعد ماشین یه چیزی رو یاد بگیره و بعد این یادگیری رو به داده های جدید تعمیم بدیم (generalize) و پیشبینی رو انجام بدیم و هدفمون هم این هست که این پیش بینی روی داده های جدید بیشترین دقت رو داشته باشه.

دو تا رویکرد برای generalization یا این تعمیم دهی وجود داره: بر اساس شباهت بین داده ها و یا یادگیری بر اساس مدل .

## Instance-Based Learning

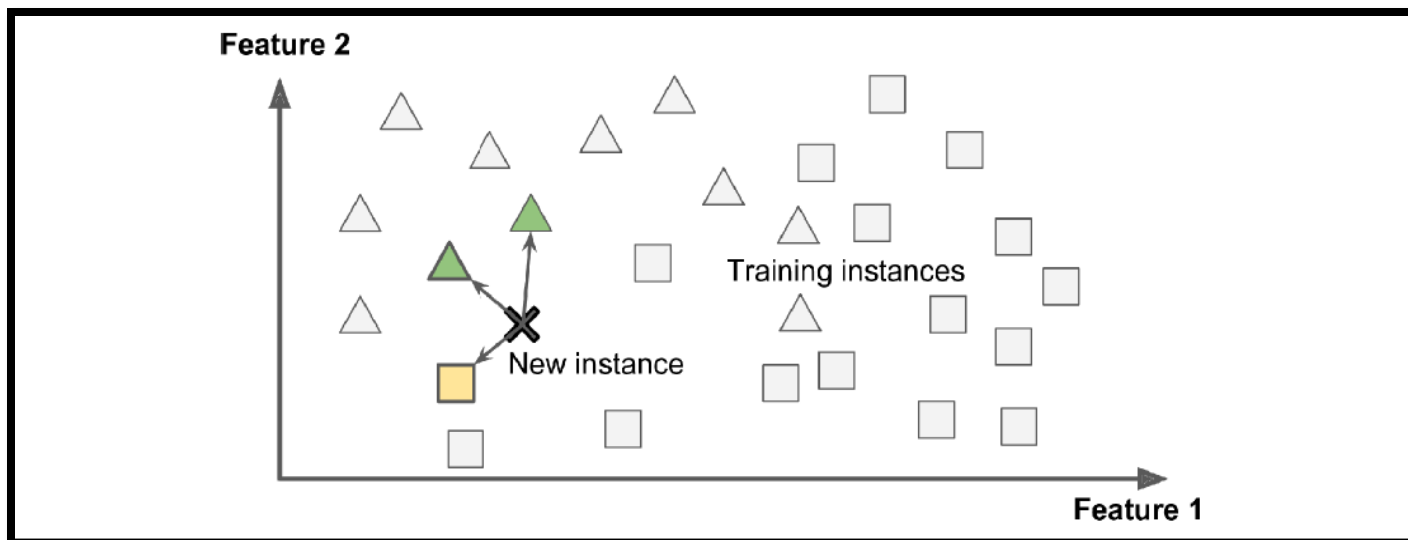
خب، یکی از بدیهی ترین نوع یادگیری میتونه این باشه که ماشین دقیقا از دیتا ها یاد بگیره و از اونا برای پیشبینی استفاده کنه. به عنوان نمونه ماشین تشخیص ایمیل اسپم رو در نظر بگیرین. ماشین ایمیل های جدید رو بر اساس مطابقت صد درصدی اون ها با داده های اولیه، پیشبینی کنه. کلمه ی "تخفیف" قبلا در ایمیل های اسپم مشاهده می شده، از این به بعد هم ماشین، هر جا با این کلمه رو به رو شد، پیشبینی می کنه که این هم یک ایمیل اسپم هست.

خب این ساده ترین حالت ممکنه اما ماشین می تونه یکم حرفه ای تر عمل کنه ، مثلا بر اساس شباهت هایی که در ایمیل های جدید و قدیم میبینه (نه لزوما برابری صد در صدی) بیاد و پیشبینی رو به ما ارایه بده. تو دلش می گه عه! این ایمیل جدید انگار بیشتر شبیه ایمیل های اسپمه تا نرمال! پس اسپم 😊



به این نوع از یادگیری ها Instance-Based Learning می گن و مدل یاد میگیره و ماشین این یادگیری رو بر اساس شباهت هایی که نسبت به هر دسته از داده ها مشاهده کرده، تعمیم می ده (generalize). تصویر پایین رو ببینین، یک نمونه جدیدی وارد سیستم میشه ، ماشین میگه عجب! این نمونه جدید دو تا شبیه مثلث هاست و یکی شبیه مربع هاست پس داده ی جدید من احتمالا از نوع \_\_\_\_ هست. (شما بگین! 😊)

معبور شدم تصویر رو بندازیم صفحه بعد 😊

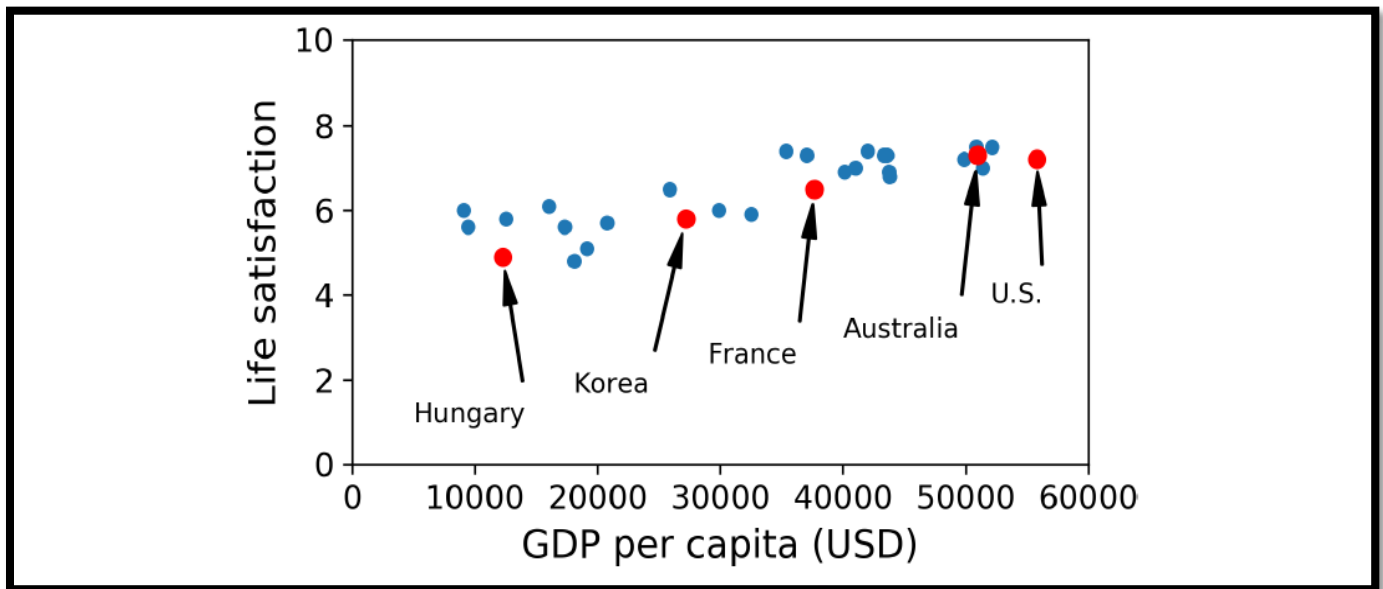


## Model-based learning

یکی دیگه از راه ها برای تعمیم دهی (generalization) ، تعمیم بر اساس مدل هست. فرض کنیم ما می خوایم بفهمیم رابطه پول با احساس رضایت در کشورها چجوریه (خب معلومه دیگه 😬) در قدم اول می گردیم دنبال یک سری داده که هم سطح رضایت و هم یک سری چیز های دیگه رو در خودش داشته باشه. جدول زیر اسامی کشور ها ، نرخ تولید ناخالصی داخلی (GDP) ها و نرخ رضایت مندی مردم اون کشور ها رو در خودش جای داده.

Country	GDP per capita (USD)	Life satisfaction
Hungary	12,240	4.9
Korea	27,195	5.8
France	37,675	6.5
Australia	50,962	7.3
United States	55,805	7.2

حالا میایم و دیتاهامون رو به تصویر در میاریم (visualization).



به نظر می رسه یه رابطه ای بین سطح رضایت و نرخ GDP وجود داره. شما هم می تونین متوجه اش بشین 🤔؟

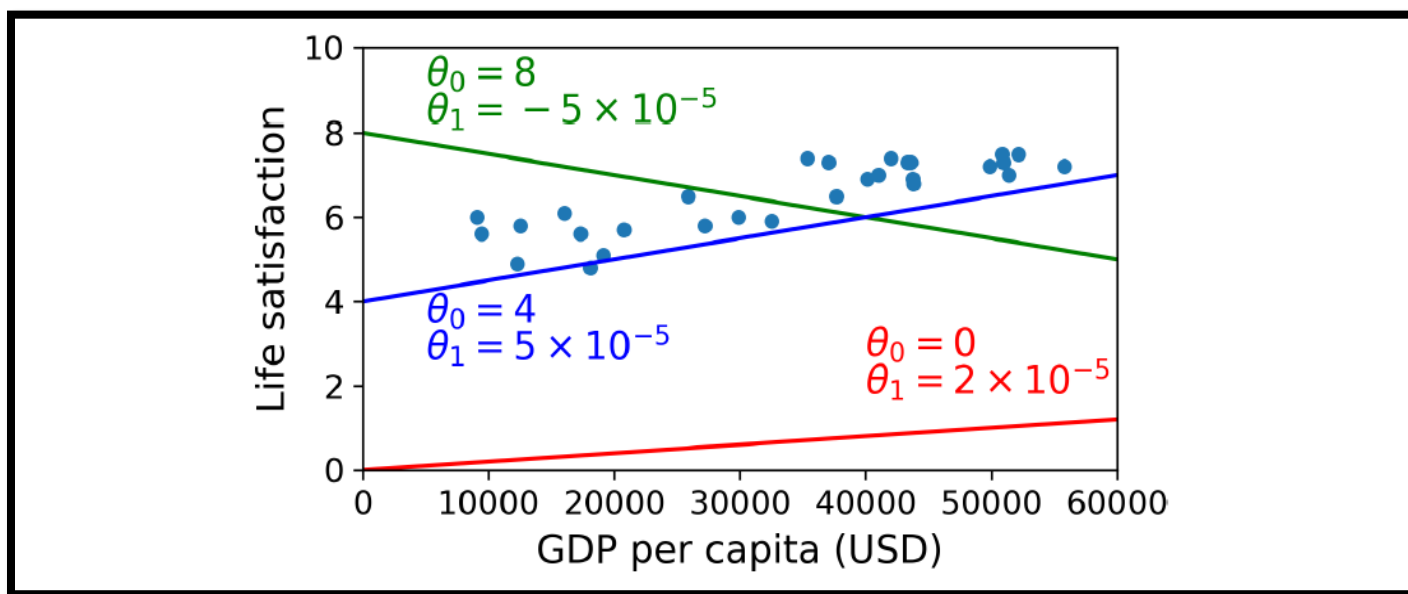
البته بالا پایین هایی وجود داره(نویز) اما اگه خیلی به جزییات اهمیت ندیم ، انگار که هر چی نرخ تولید ناخالص داخلی بیشتر می شه، سطح رضایت هم رشد میکنه. اینجاست که یک تابع رو برای ماشینمون انتخاب می کنیم. به این مرحله (**model selection**) می گیم. مدل می تونه از جنس و شکل هر تابعی باشه. ظاهرا اینجا یک مدل خطی (linear model) می تونه بهمون خروجی مناسبی رو بده.

$$\text{life\_satisfaction} = \theta_0 + \theta_1 \times \text{GDP\_per\_capita}$$

خب نوع مدلون رو تعیین کردیم اما چه ضرایبی باید برای اون تعیین کنیم تا مناسب ترین خروجی بهمون تحویل داده بشه. تتا ( $\theta$ ) معمولا برای نمایش پارامترها (ضرایب) مدل به کار میره.

در اینجا مدل ما دو تا پارامتر داره.  $\theta_0$  و  $\theta_1$

به ازای مقادیر مختلف این پارامترها، پیشبینی ماشین برای داده های جدید، می تونه متفاوت باشه. در تصویر زیر هر خط دارای پارامترهای متفاوتی هست. تتاهای متفاوت، خطوط متفاوت!

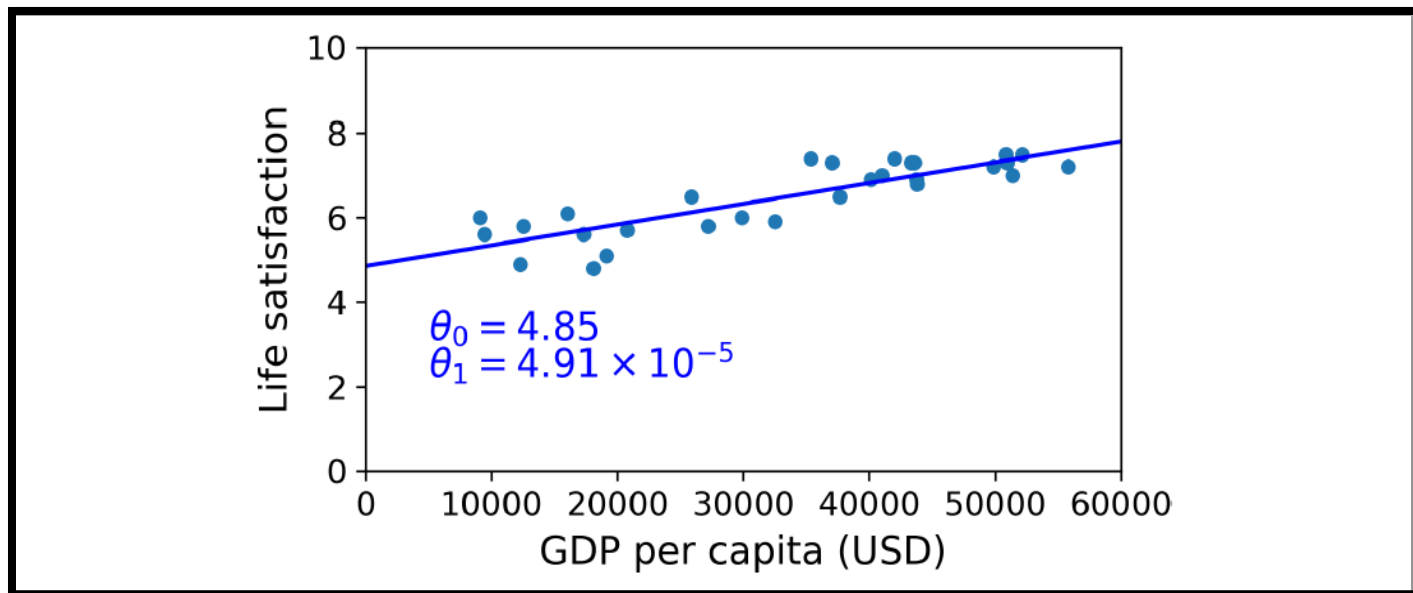


اما سوال اصلی اینجااست. چجوری بهترین پارامترها رو برای مدلون تعیین کنیم؟ برای این کار باید برای ماشین، یک شاخص عملکردی (performance measure) تعریف کنیم.

معمولا برای مدل های خطی یک تابع هزینه (cost function) تعریف میکنن که بیانگر تفاوت مقادیر پیشبینی شده با مقادیر واقعی هست در مقیاس های متفاوت هست. هر چی این مقدار کمتر باشه مدلون بهتره و ماشین سعی می کنه این شاخص رو به بهترین (کمترین) حالت خودش برسونه.

نتیجه ی بهترین حالت، یک مدل با پارامترهای مشخص و بهینه برای شماست. 😊

به این پروسه یعنی آموزش نمونه های اولیه به ماشین و یافتن بهترین پارامترها برای بهترین سازگاری با دیتاها، آموزش مدل گفته می شه (training the model)



خب به سلامتی الان یه ماشین هوشمند داریم! فرض کنیم می خوایم نرخ رضایت کشور قبرس (Cyprus's) رو به دست بیاریم. نرخ تولید ناخالص داخلی برای این کشور ۲۲,۵۸۷ هست. با توجه به مدل و پارامترهایی که ماشین بهشون رسیده، این پیشبینی به صورت زیر به دست میاد. و تمام!

$$4.85 + 22,587 \times 4.91 \times 10^{-5} = 5.96.$$

حالا اگه رویکرد یادگیری ماشینمون از نوع (Instance-Based Learning) بود، ماشین میومد و نزدیک ترین کشورها از نظر نرخ تولید ناخالص ملی به قبرس رو پیدا می کرد و نرخ رضایت در اول کشور رو مشاهده! و در نهایت نرخ رضایت در کشور قبرس رو میانگینی از نرخ رضایت در کشورهای

نزدیک به اون، پیشبینی می کرد. به این روش نزدیک ترین همسایه (K-Nearest Neighbors) گفته می شه که بعدا حتما بهش می رسیم.

اگه همه چی خیلی خوب پیش بره، ماشین به شما یک پیشبینی دقیق رو ارایه میده. اما در غیر این صورت شاید نیاز باشه یک مدل دیگه برای ماشین در نظر بگیرین یا شاید نیاز به فیچر هایی علاوه با GDP مانند جمعیت، آلودگی هوا، نرخ بیکاری و ... داشته باشین؛ یا حتی اصلا باید داده های بیشتری به مدل بخورونین! شایدم دیتا هایی که به مدل خورونده بودیم! دیتا های به درد نخوری بودن. توجه به این جزییات که ما رو تبدیل به یه دیتا سانیس حاذق می کنه. 🐼

تا اینجا کار فهمیدیم یادگیری ماشین چی هست، چی نیست!، به چه کاری میاد اصلا!، چه دسته بندی هایی وجود داره، روند کلی کار به چه صورت هست و این که اگر تحریم برداشته بشه و کارخونه ها رونق بیشتری بگیرن و GDP مون رشد کنه، احتمالا نرخ رضایت ما ایرانی ها هم بیشتر میشه 😞.

در قسمت بعدی به این می پردازیم که در این مسیر پر پیچ و خم چه چالش هایی می تونه وجود داشته باشه و چه چیز هایی می تونه از ارایه یک پیشبینی خوب توسط ماشینمون جلوگیری کنه.

پس با ما همراه باشه.

راستی این وسط ها اگر خواستین از ما حمایت کنین و برای ادامه کار بهمون انگیزه بدین، فقط کافی این مطالب رو با دوستای دیگه تون، به اشتراک بزارین.

تا درودی دیگه، بدرود! 😊\