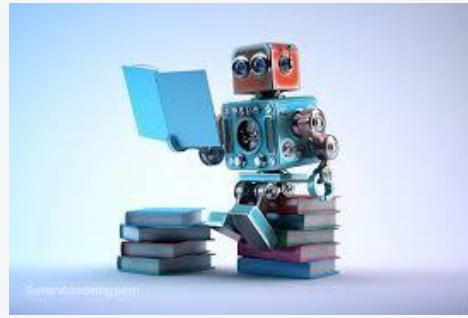


به نام خدا



یادگیری ماشین





یادگیری ماشین

آرش عبدی هجراندوست

arash.abdi.hejrandoost@gmail.com

دانشگاه علم و صنعت
دانشکده مهندسی کامپیوتر
نیم سال اول ۱۴۰۱-۱۴۰۲

در این جلسه

✗ مقدماتی از یادگیری ماشین

✗ انواع یادگیری

✗ یادگیری با نظارت

یادگیری

❌ یکی از بهترین ویژگی های یک ماشین هوشمند

❌ تقلید از موجودات هوشمند طبیعت

○ شکل گیری دسته بندی در کودکی با دیدن نمونه

❌ کاربردی شدن بیشتر هوش مصنوعی

○ مجموعه ای از داده ها

○ الگوریتم یادگیری

○ پیش بینی خروجی

سیستم یادگیرنده

❌ چیست؟

- آیا اگر نرم افزاری هوشمند، در طول زمان استفاده، عملکرد خود را بهبود ندهد، می‌توان آن را یادگیرنده دانست؟
- معنای عرفی؟ علمی؟

سیستم یادگیرنده

✗ یک سیستم یادگیرنده است اگر کارایی خود را به مرور زمان و با مشاهدات جدید بهبود دهد.

- از افزایش کارایی در سرعت نوشتن یک شماره تلفن (یا زدن امضا)
- تا ارائه یک نظریه جدید درباره جهان توسط انیشتین

✗ یک نوع پرترفدار از یادگیری:
○ یادگیری تابعی که خروجی را تخمین میزند، با داشتن مجموعه ای از جفتهای ورودی-خروجی

لزوم یادگیری

✗ سوال: چرا عامل یاد بگیرد؟

○ چرا خود طراح عامل هر آنچه لازم است را ابتدائاً به خود عامل ندهد؟

✗ طراح ممکن است همه حالتها را نداند

✗ طراح ممکن است همه تغیرات محیطی در طول زمان را نداند

✗ طراح خیلی وقتها نمی‌داند که یک چیز را چگونه می‌داند، بنابراین

نمی‌تواند برنامه‌اش را بنویسد:

○ شناخت اعضای خانواده از روی تصویر چهره

یادگیری : چه چیز را ؟

✗ خودروی هوشمند:

○ کی ترمز کند؟ آنچه می بیند کی اتوبوس است؟ تاثیر یک عمل در جاده فیس چیست؟ وقتی در آخر روز پول چندان از مسافران گیرش نیامد، بفهمد که کلا کارا نبوده...

✗ دانش اولیه؟

✗ نحوه بازنمایی دانش

✗ بازفورد موجود از محیط برای بهبود

برخی کاربردهای موفق یادگیری ماشین

✗ شناسائی الگو:

- شناسائی چهره و حالات آن
- شناسائی حروف دست نویس
- شناسائی گفتار

✗ شناسائی رفتار های نادرست

- تشخیص خرابی سیگنالهای سنسور ها
- تشخیص سوء استفاده از کارت های اعتباری

✗ پیش بینی

- قیمت سهام
- پیش بینی قیمت ارز

یادگیری استقرایی (INDUCTIVE)

✗ ورودی ها برداری از مقادیر ویژگی (Attribute) است

✗ خروجی ها : مقدار پیوسته یا گسسته

✗ یادگیری از از روی جفت نمونه های ورودی - خروجی یادگیری استقرایی

نام دارد

○ رویکرد bottom-up

✗ یادگیری استنتاجی (Deductive) در مقابل یادگیری استقرایی است که

در آن با کمک دانش و قواعد موجود، دانش جدید کشف میشود.

○ رویکرد top-down

بازخورد در یادگیری

✗ یادگیری غیرنظارتی (Unsupervised)

○ یادگیری الگو هایی از ورودی بدون دانش بازخورد (خروجی)

○ فوشه بندی

○ مثلاً خودروی هوشمند ممکن است کم کم مفهومی به نام «روز خوب ترافیکی» یا «روز بد ترافیکی» را یادبگیرد بدون آنکه از کسی بازخوردی گرفته باشد (صرفاً با مسگرها و ادراکات ورودی خودش)

بازخورد در یادگیری

✗ یادگیری تقویتی (Reinforcement)
○ بازخورد = پاداش یا جریمه

■ برای هر عمل یا برای مجموعه ای از اعمال

■ تصادف در اعمال خودروی هوشمند

بازخورد در یادگیری

✗ یادگیری با نظارت (Supervised)

○ داشتن نمونه‌هایی از ورودی-خروجی و یادگرفتن تابعی که ورودی را به خروجی نگاشت کند.

- مجموعه‌ای از تصاویر با تگ اتوبوس/ناتوبوس
- مجموعه‌ای از ادراکات خودرو با تگ «ترمز کن» یا «پیچ به چپ» که توسط انسان داده شده.

بازخورد در یادگیری

✗ یادگیری نیمه نظارتی (Semi-Supervised)

- داشتن تعدادی داده با برچسب خروجی و تعدادی بدون خروجی
- متی خروجی ها ممکن است دقیق یا درست نباشند (نویز)
 - تخمین سن بر اساس چهره
 - دروغ گفتن سن توسط اشخاص یا نا دقیق گفتن آن

✗ در هر دو صورت نویزی بودن خروجی یا فقدان خروجی، چیزی بین یادگیری با و بی نظارت داریم.

یادگیری با نظارت

✗ مجموعه ای آموزشی از N نمونه داده با فرمت ورودی-خروجی به شکل زیر داریم:

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

که در آن مقادیر y_i بر اساس تابعی مجهول به شکل $y = f(x)$ تولید شده اند. وظیفه یادگیری با نظارت یافتن تابع h است که مقدار تابع صحیح (یعنی f) را تقریب می‌زند.

✗ در اینجا x و y میتوانند هر مقداری (حتی غیر عددی) داشته باشند.

✗ y_i را ground truth (جواب درست) می‌نامیم.

یادگیری با نظارت

- ✗ تابع h یک فرضیه (Hypothesis) نام دارد.
- ✗ یادگیری وظیفه دارد در بین فرضیه های **ممکن**، بهترین آنها را

بیابد

- بهترین در داده های آموزشی
- و داده های آزمایشی

✗ تعابیر مشابه

- h یک مدل از داده ها است.
- h یک تابع از داده ها است.

دسته‌بندی و تقریب

✗ اگر خروجی y مقداری از یک مجموعه محدود داشته باشد (مثلا بارانی-ابری-آفتابی)، مساله یادگیری، یک دسته بندی نام خواهد داشت

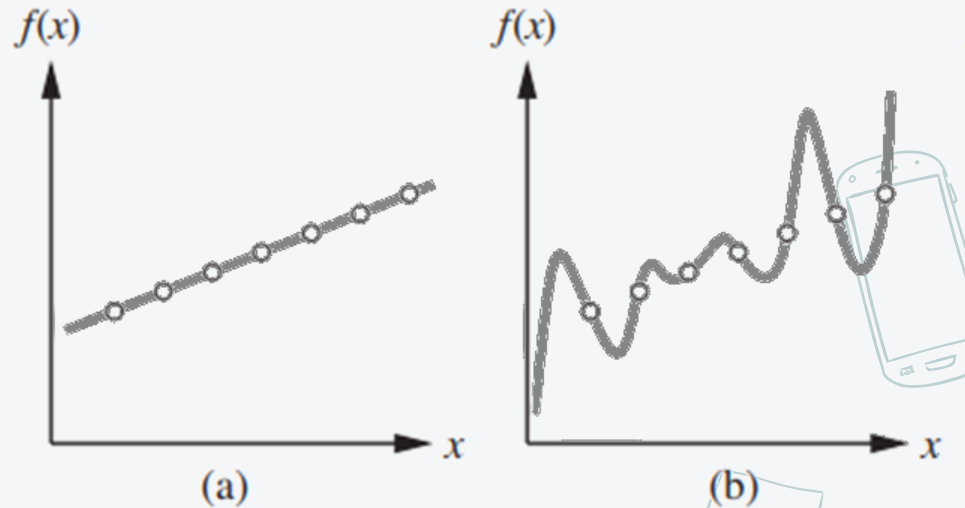
○ و اگر تنها دو مقدار خروجی ممکن باشد، دسته بندی دودویی (Boolean-Binary)

✗ اگر خروجی y یک عدد (نامحدود) باشد (مثلا دمای فردا)، مساله یادگیری، Regression نام خواهد داشت.

دست‌بندی و تقریب

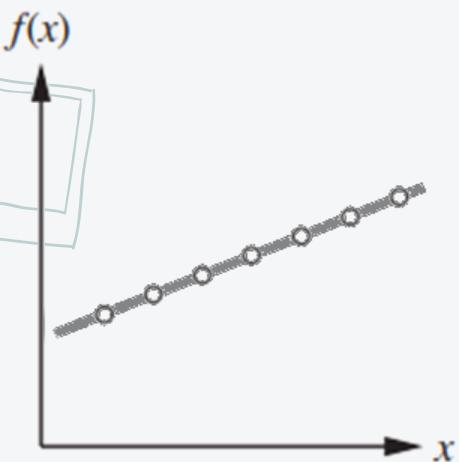
- ✗ به طور دقیق، رگرسیون شامل یافتن امیدریاضی یا متوسط مقدار y (برای هر ورودی) می‌باشد، زیرا احتمال یافتن مقدار دقیق خروجی صفر است.
- آنچه مشاهده شده، صرفاً تعدادی نمونه است و با استقرا به دنبال فهمیدن حقیقت هستیم.
 - حتی اگر مقدار خروجی برای یک ورودی معین x در همه داده‌های مشاهده شده، دقیقاً یک عدد y باشد، امیدریاضی خروجی برای x ، y است، اما احتمال آنکه خروجی برای ورودی x بعدی (مشاهده نشده) دقیقاً y باشد، صفر است! (مفهوم حد)
 - احتمال آنکه خروجی، مقداری در یک بازه داشته باشد، میتواند غیر صفر باشد، اما احتمال عدد دقیق برای خروجی صفر (حدی) است.

نمونه های از رگرسیون

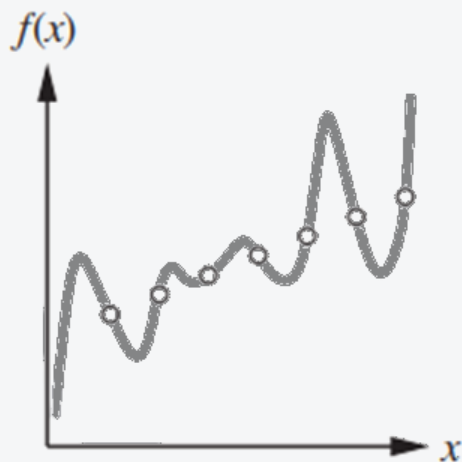


انتخاب تابع h از فضای فرضیه های H ❌
 فرضیه سازگار: با نمونه های موجود منطبق باشد ❌

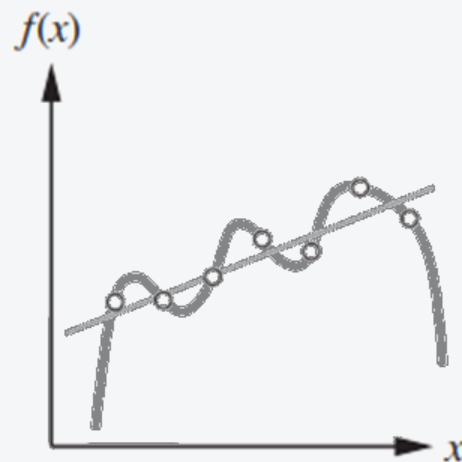
نمونه های از رگرسیون



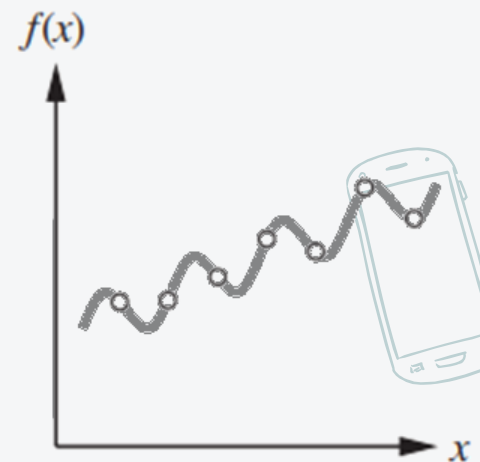
(a)



(b)



(c)



(d)

انتخاب تابع h از فضای فرضیه های H ❌
 فرضیه سازگار: با نمونه های موجود منطبق باشد ❌

انتخاب فرضیه

✗ انتخاب از بین چند فرضیه سازگار؟ (در H)

✗ برقراری توازن بین:

○ سادگی ~ تعمیم

○ پیچیدگی ~ دقت

✗ توسعه فضای فرضیه H به سینوسی در شکل قبل (بخش d)

○ تامین سادگی (تعداد پارامتر کم) و دقت، با هم

✗ اهمیت فضای فرضیه

✗ یک مساله یادگیری امکان پذیر یا شدنی است اگر فضای

فرضیه شامل تابع درست باشد.

انتخاب فرضیه در یادگیری با نظارت

✗ با داشتن نمونه های داده (data)، انتخاب فرضیه در یادگیری با نظارت به شکل زیر قابل بیان است:

$$h^* = \operatorname{argmax}_{h \in \mathcal{H}} P(h|data)$$

$$h^* = \operatorname{argmax}_{h \in \mathcal{H}} P(data|h) P(h)$$

✗ با این فرمول، میتوان گفت احتمال $P(h)$ برای تابع چند جمله ای درجه ۱ یا ۲، بالا و برای چندجمله ای درجه های بالاتر، کم است.

✗ با دادن احتمال پایین برای چند جمله ای های درجه بالا، اجازه می دهیم فروجی یادگیری تابعی پیچیده باشد، وقتی که داده ها واقعا نیاز به چنان تابع پیچیده ای داشته باشند.

✗ در واقع بخش $P(h)$ برای توابع پیچیده نقش جریمه را بازی می کند.

فضای فرضیه

- ✗ با توجه به اهمیت فضای فرضیه، چرا فضای فرضیه را به تمامی فرضیه های ممکن توسعه ندهیم؟
- تمامی فرضیه های ممکن: تمام برنامه های کامپیوتری قابل نوشتن یا یک ماشین تورینگ
 - مطمئن باشیم فرضیه درست در مجموعه H وجود دارد

فضای فرضیه

× پیچیدگی یافتن فرضیه درست

× برقراری توازن بین:

○ میزان رسا بودن و منطبق بودن بودن فضای فرضیه با داده ها
(Expressiveness)

○ پیچیدگی/سادگی یافتن یک فرضیه در آن فضا

× فرضیه ساده تر:

○ یافتن ساده تر

○ استفاده کم هزینه تر

فضای فرضیه

✗ برقراری توازن بین پیچیدگی/سادگی فرضیه و میزان رسا بودن فضای فرضیه، چندان ساده نیست:

○ گاهی فضای فرضیه رسا می‌تواند سبب یافتن یک فرضیه ساده (و در عین حال، سازگار) نیز بشود.

○ گاهی محدود کردن میزان رسا بودن فضا، سبب می‌شود هر فرضیه سازگاری لزوماً پیچیده باشد

✗ مثلاً، حذف \sin از فضای فرضیه (محدود کردن میزان رسا بودن)، سبب می‌شود فرضیه چندجمله‌ای به شکلی بسیار پیچیده (درجه خیلی بالا) پیدا شود (با فرض آنکه سازگاری آن ممکن باشد)

انتخاب فضای فرضیه

✗ نیازمند دانش پیشین از دنیای تولید کننده داده‌ها

✗ در صورت فقدان چنین دانشی

○ آنالیز اکتشافی داده‌ها

○ انجام آزمون‌های آماری

○ نمایش آماری داده‌ها

■ هیستوگرام‌ها

■ انواع نمودارها (نقطه‌ای، جعبه‌ای، ...)

✗ باید سعی نسبت به داده‌ها پیدا کرد تا بتوان فضای فرضیه را مشخص نمود.

✗ وگرنه، باید صرفاً چند فضای فرضیه را در نظر گرفت و آزمود!

آگاهانه

انتخاب فضای فرضیه

✗ نیازمند دانش پیشین از دنیای تولید کننده داده‌ها

✗ در صورت فقدان چنین دانشی

کارشناس ارشد هوش مصنوعی

یا فوق دیپلم شبکه‌های عصبی؟

○ آنالیز اکتشافی داده‌ها

○ انجام آزمون‌های آماری

○ نمایش آماری داده‌ها

■ هیستوگرام‌ها

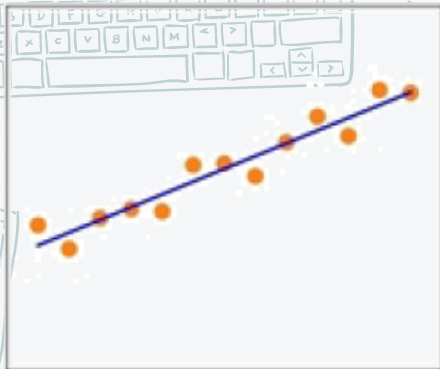
■ انواع نمودارها (نقطه‌ای، جعبه‌ای، ...)

✗ باید سعی نسبت به داده‌ها پیدا کرد تا بتوان فضای فرضیه را مشخص نمود.

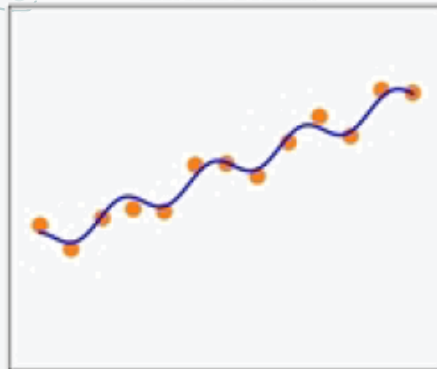
✗ وگرنه، باید صرفاً چند فضای فرضیه را در نظر گرفت و آزمود!

Linear

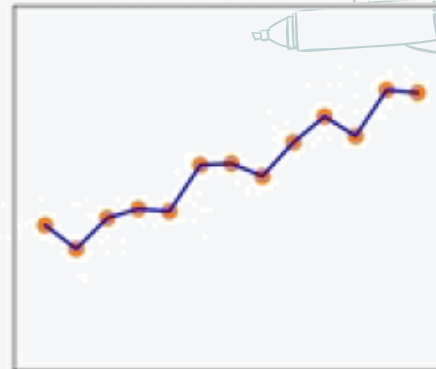
Data set 1



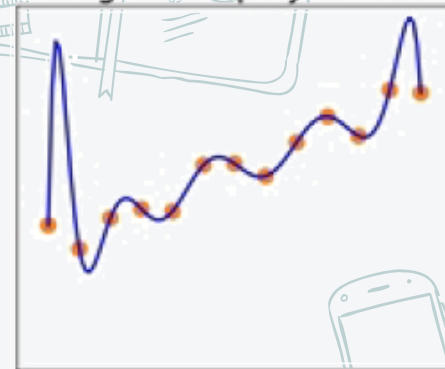
Sinusoidal



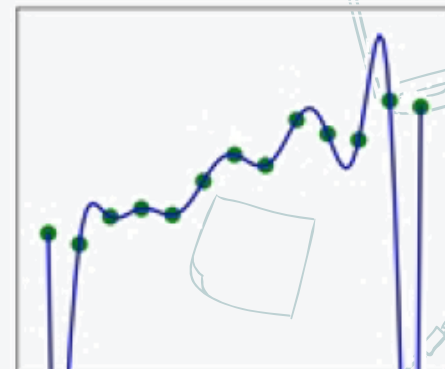
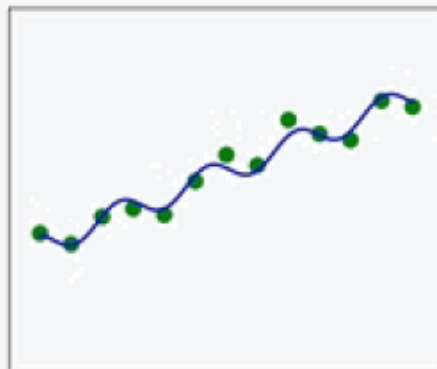
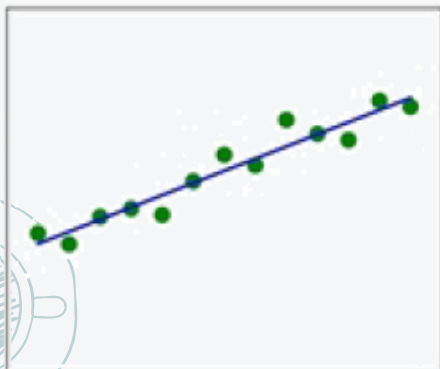
Piecewise linear



Degree-12 polynomial



Data set 2



✗ انطباق فرضیه از فضاهای مختلف فرضیه روی دو مجموعه کمی متفاوت از داده‌ها
 ○ هر دو مجموعه از یک تابع یکسان نمونه‌برداری شده‌اند

BIAS-VARIANCE

✗ یک راه برای تحلیل فضای فرضیه

✗ Bias:

- میزان گرایش فرضیه/مدل برای منحرف شدن از مقدار درست در dataset های مختلف
- محدود بودن فضای فرضیه
- مثلاً نمودار فنی در مثال گذشته
- Underfitting
- پیدا نشدن الگوی مناسب در داده‌ها (ی آموزشی)

✗ Variance:

- میزان تغییر فرضیه هنگام تغییر dataset
- Overfitting

BIAS-VARIANCE TRADEOFF

✗ انتخاب ساده‌ترین فرضیه که با داده‌ها سازگار باشد.

✗ پارامتر بیشتر ؟ = تعمیم کمتر

○ نه همیشه؟

○ شبکه‌های عمیق

■ میلیون‌ها پارامتر

■ تعمیم کافی

■ پارامتر بیشتر :

● بیان دقیق‌تر واقعیت پیچیده موجود.

● تنبلی و مفا کردن داده‌ها



خسته نباشیم!

با تشکر

