



# تمرین های درس مبانی داده و پردازش



پخش ۱ یا دستگیری ماشین =

سوال یک =  $\text{Supervised Learning}$  و  $\text{Unsupervised Learning}$

چه تفاوتی دارند؟

بادلگیری با دظارت معمولاً در زمینه دسته بندی به کار گرفته می شود و فقط بخراهم

ورودی را در بر چسب های دسته خروجی با خروجی پرسنلیتی داشته اند، این نوع  
بادلگیری به نام مارخ می آید. از جمله الگوریتم های متداول در بادلگیری با دظارت می توان

برگرسیون لجستیک، بینی ساده و ماشین های بردار پشتیبان (SVM) شبکه های مصنوعی  
عصبی و غیره اشاره کرد.

از جمله متداول ترین کارهای این نوع توان بادلگیری برون استفاده انجام داده می شوند  
تران به خوش بندی بادلگیری اراده در تخصیص چنان اشاره کرد. در عین موارد بروزگرانی بادلگیری  
ساختار دقیق داده برون استفاده از برچسب های داده هستیم. برخی از الگوریتم های رایج عبارت

اندازه نمودن های  $K$  means  $K$  تحلیل مولنده اصلی خود را نمایند. چون همچویی برچسبی  
ارائه نشود، همچویی مشخصی برای متایسه محیط اسلامی در اثر روش های بادلگیری برون  
نمایه است وجود نیست. روش های بادلگیری برون نظارت در تحلیل داده های آرثناهی و  
کمتر بعد نیز مردم استفاده فراهم گشته

در مادلیهای نظارت شده، الگوریتم از جویه داده های آموزش با پیش بینی همراه

و رئالیسم برای تولید پاسخ صحیح کمالیگری و حاکم به صد های یادگیری نجت نظارت نسبت به صد های یادگیری بون نظارت دقیق تر است، اما برای برچسب لذاری مناسب داده های مدخله اولیه انسان نیاز دارد.

در مداخله که صد های یادگیری بون نظارت که توانی برای رشف ساختار ذاتی

داده های بون برچسب کاری نهفته، البته باید توجه داشت که آنها مسزدم هنر یا حریم از مداخله انسانی برای متغیرهای خروجی نیاز دارند.

چرا وظایف Machine Learning مای الگوریتم ضروری است؟

هدف از مقایسه بینی وینگرک این است که امیدواران حاصل شود که همه وینگرک ها بون

و نظرگیرنده مقایسه اصلی شان، به حدودیکسان (نمایندگان یادگیری مشاهدات) نشوند.

در مدل یادگیری ماشین مانع خواهیم وینگرک خاصی تاثیر پیشتری از بقیه داشته باشد.

و ما باید وینگرک هارا بگردانیم مثلاً این کنیم که یک وینگرک با بزرگی زیاد بر سایر وینگرک ها تسلط پیدا نماید.

C. تناوی Normalization و standadization چیست

هنگام سازی زمانی ضریب انتشار تردد داده های شما از تردد بیکاری پیروی نمی کند. این می تواند الگوریتم های ضریب باشد که سیچ تر فیکی از داده های امراض نمی کند، همانند

۱- نزدیکتری مسایی با شبکه های عصبی، از طرف دیگر، مقایسه سازی می تواند در میان

ضریب باشد که داده های از تردد بیکاری پیروی می کند. با این حال، لازم نیست که لین امیزها

درست نباشد) اما اگر توزیع دیگری شما گزینی باشند، این نکته معتبر است. همچنین بر خلاف هنجارسازی محدود محدودیت فذارده بنا بر این، حتی اگر در داده های خود

متغیر زیادی داده دورانه داشته باشند، تحت تاثیر معادف سازی ضرایب خواهد گرفت. با این حال، انتخاب استناده از هنجارسازی یا صفاتی سازی بستگی به ساختار شما و الگوریتم داده هایگری م Ashton شمادارد. هیچ قانون سخت وجود ندارد که به شاگردیدن چه زمان داده های خود را صفاتی یا هنجارکننده، شما سیمه هی توانید مدل خود را با داده های خام، هنجار و مصادر پنهان شده تراویح نمایی و عملکرد را برای بهترین نتایج مناسب بگینید.

## D - چرا متناسب بینی Min-Max Normalization

داده های استناده می شوند؟

در این نکته، داده های یک معروفه متشخص از نظر می باشند،  $x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$  یک روش برای مقادیر بینی داده هاست. پس از انجام  $\text{Scaling}$ ، مقدار داده به بازه ۰ تا ۱ نگاشت می شود. روش انجام این نرم افزار سازی سه پساده اند. کافی است ابتدا مقدار مینیمم و مکزیمم داده را بپرسیم. با استفاده از فرمول زیر، داده را متناسب بینی

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

$x'$  = مقدار داده

$x_{\min}$  = مقدار مینیمم داده

این روش مناسب داده های است. که توزیع نرم افزار و نیاز به متناسب بینی در بازه خاص دارند. - مثلاً با این مشخص بوده و دیگر ناقص باشند تواند

نموده پرست می باشند / وقتی توزیع داده های متشخص نیست - یا ای داده های موزیس به جزء

دستگاه دسی) می باشد و متریک پکنداخت ایضاً

Z-score - Normalizath - E

در دوبن نرسان سازی Z-score مبتنی بیم داده گالام مرض شده است و از مین هیت که بیشتر سایر داده های می باشد (متغیرهای میانگین) برای رایه (انحراف میانگین) باشد در فرمول زیر نماید  $\text{Z-score} = \frac{x - \mu}{\sigma}$  بترتیب بیانگیر میانگین و انحراف میانگین است:

$$z_{\text{standardized}} = \frac{x - \mu}{\sigma}$$

به تاریخی این تکنیک به وزیره زمائن صنایع است که الگوریتم یادگیری مانند مدل های خطی، ترجیح داده های از نرمال در خود پذیرد. برخلاف متغیر میانگین کمینه - پیشیگیری کامل است اما مدل های خطی محدود به دامنه خاصی نیست.

Machine Learning  $\rightarrow$  Regularization  $\in F$

ب چه صورت است؟

منظم سازی (Regularization) تکنیکی است که برای جلوگیری از مشکل بیش برداشت استفاده می شود. منظم سازی روی نهاد مدل های یادگیری ماشین خابل پیاده سازی است. ذکر نموده تکنیک منظم سازی به لین صورت است که با اضافه کردن ضریب ثابتی به تابع  $J(\theta)$  از پیشگیری مدل های که اختلال رخداد پیش برداشت در آنها برداشت باشند کاسته می شود. مدل که پیش برداشت باشند بین نتایج پیش بینی دقیقی از نمونه درودی های جبریده را دارد. دسر با اصطلاح با مثیل (عمی سازی) دراجه است.

منظم سازی 1 و منظم سازی 2 در سوی رایج منظم سازی هستند.

←

## under fitting & over fitting

### Machine building بی جودتی آوردن؟

کم برآش (Underfitting) سناریوی در حلم داده یا دیتا ساینس است که در آن یک مدل قادر نیست رابطه میان متغیرهای ورودی و خروجی را به طور دقیق ثبت کند.

این خطایجا دخالتی با لایس در هر دو مجموعه آزمایش و مجموعه داده های دبره تشدید را رقمی زند.

این اسرارمانی رخ می سد که مدل بیش از حد ساده باشد. بین مدل‌ها مدل پر زمان آمدن بیشتر ویژگی های ورودی پیشتر یا تنظیم (Regularization) متر نیاز دارد. مدل در لس برآش نمی تواند الگوی (نکته) داده ها را تشخیص دهد. در نتیجه این امر افلاطی خطا و عملکرد فسیف مدل را به همراه دارد.

اگر مدل نتواند بخوبی به داده های جدید تعیین داده شود، نمی تواند از آن برای طبقه بندی یا پیش بینی استفاده کرد. تعمیم یک مدل به داده های جدید در نتیجه همان چیزی است که به ما امکان می دهد سری دلخواهی را می توانیم ایجاد کنیم. مدل هایی که می توانند این امکان را داشته باشند مدل های افتخاری می نامیم.

مفهوم Over fitting در باگیری مانع در زبان فارسی بانام بیش برآش خوانده می شود و زمانی احتراق می افتد که مدل ماسن می کند نه از نقاط طده یا بیشتر از نقاط داده مورد نیاز می گردد در مجموعه داده را پوشش دهد و همین دلیل مدل شرمند یا ذخیره نمی نماید. مدل دیر که دست بیشتر در مجموعه داده در مجموعه داده می کند و همین این مطالعه کارایی و دقت آنرا کا هش می دهد. بیش بیش بزرگ دارایی با یا کم واریاس

پالاست

احتمال بروز پیش هرآنچه برآندازه ای رئیس مدل خود آموزش می سیم افزایش  
می یابد و این بران مناسب است که مرحله بیشتر مدل خود را آموزش دیگر کاشش بیشتری بدل  
رخ دادن پیش هرآنچه بوجود می آید Over fitting یعنی از مسئله اصلی  
در یادگیری تحت نظر است.

Train / Test Split / Cross - Validation - H

کاربرد دارد؟

اعتبارسنجی متناسب تکنیک برای ارزیابی یک مدل یادگیری ماشین را زی پیش  
مکنند آن است CV معنواً دلایل داده کاربردی ML استفاده می شود  
CV لبکه می کند تا مدل های مختلف را توانیم و مناسب ترین را با عرضه به  
مستله پیش بین خود انتخاب کنیم.

دک CV آسان است پیاده سازی آن آسان است و با پاس کسری  
ذیت بسایر دوشنایی مزده استارده برای تحسین امتیاز نهادی مدل دارد و هم  
اینها اعتبارسنجی متناسب را به این مدل تقدیر نموده اند این مدل بهترین  
مدل برای نادر خاص تبدیل می کند.

تلنگ مان مختلف وجود دارد که مکنن است برای اعتبارسنجی یک مدل استناده  
شود با این حال همه آنایک الگوریتم مشابه دارند.

۱. تقدیم دیناست بد در حقیقت آموزش و تست

۲. آموزش مدل با استناده از داده های آموزش

clips

۳. تکرار چند باره مراحل

## Hold-out روش

اعتلار سنج متنامه ( $H_0$ -out) ساده ترین دایجی ترین تگلیل است. شاید نداشته باشد که اسم آن را همچنان  $(H_0$ -out) نامیدیم.

۱- دیتاست را بدوبخش آمرزش و تست تقسیم کنید، محمدگاه در حدیات  
برای آمرزش ده درصد آن برای دست کنارگذاشته من شود، البته شما هم ترانیده مرتدم  
بنده را (بندخانه) لئن نه برای بسیار مناسب تراست.

2- مدل رایا (ستاده‌ای ابداده‌ای آموزشی (ai/aɪ) آموزش می‌دسم.

۳. مدل داده‌های ارزیابی Validation (ارزیابی می‌کنیم).

## 4 نتائج Validation را ذخیره کنیم.

ما معمولاً از روش **Hold-out** برای دیتاست های بزرگ استاده می کنیم،  
زیرا این روش نتایج معتبر نیاز به آزمون مدل دارد.

Gradient Descent

گریدیان کامپت لاین بلکه الگوریتم بهینه‌سازی تکراری سرتیبه اول برای یافتن  $\text{Gradient Descent}$  کمینه‌های موضعی را نیز یکه تابع مشتق پذیر است.

می دانیم که برای یافتن گمینه مای که تابع دامنه‌اللئ تر بخواهیم گلوبیم برای

لایفتن کالسترم سایلیت تابع می توانیم مشترک آنرا بگیریم و مساوی صفر مراد دسیمه جواب

میں چین معاشری نظری اذتابی راخراہنداد کے تابع متدار minimum

دارد و دیشنه) (لیکن maximum ک

ج = <

چرا Deep Learning برای پیچیده ترین مسائل استفاده می شود؟

باید گیری عمیق یک زیر شاخه از بادگیری ماشین و بر مبنای مجموعه ای از  
الگوریتم هاست که در تلاشند مقامی انتزاعی سطح بالا در دادگان مدل کنند که این  
فرایندهای استاده (ند) یک گراف عمیق که دارای چندین نایه پردازش تعلمی  
متصل از چندین نایه تبدیلات خطی و غیرخطی اند که مدل می کنند به بیان دلیر  
نمایند آن باید گیری نایش داشت و دویگی های دلایلی مدل است.

به قولی دلگیر «باید گیری عمیق» یک نوعی «باید گیری ماشین» به همراه تسلیمهای  
عمیق چندلایه است که با تلقی فرایندهای الگوریتمی موجود در داده ها کشف کرده و  
به دستور خود از آنها تواند علاوه بر این اشیا را شناسایی کرده و زبان مادری پیغام

|        |
|--------|
| دانشجو |
| فرمیدن |
| ضدراحت |

پیش (روم)

|             |
|-------------|
| دورس        |
| محمد        |
| احمد فرزاده |

Python

Programming

A - هر Python زبان برنامه نویس محبوب علم داده است.

پایتون ویژگی های منحصر به فردی دارد که باعث می شود استاده از آن در محاسبات کمی و تحلیلی آسان باشد. این زبان برنامه نویس برای مدت طولانی بعنوان پیترس رو داده صفت شناخته می شود و به طور گسترده در حوزه های مختلف کامپیوتری و گانه برداشته سیگنال، کامپیوچر، دیگر، استفاده می شود.

ذمان پایتون به طور گسترده در حوزه دیناساینس (Data Science) یا صناعتی علوم داده استاده می شود و یک ابزار صردد عدایته برنامه نویسان است. که علامه هر این پایتون یک زبان منعطف و متبع باشد. لذا بخانه های خفیه آن برای دست کاری داده ها استفاده می شود. با دلیلی برای یک تحلیلگر مبنی داده بسیار آسان است.

چه اینه پایتون یک پلتفرم مستقل است که برای تجزیه با مرزی رساخت موجود که می تواند برای حل پیچیده ترین مشکلات استفاده شرده اگر باشد ما از آن برای تجزیه و تحلیل داده ها موسسات برای تجسم داده د پردازش داده و شرکت های پیش بینی آب و سوا به مانند تجزیه و تحلیل Forecasting و Watch از آن استفاده می کنند.

- علت های ترجیح داده شدن پایتون:

انتخاب کتابخانه ها

قدرتمند و آسان برای استفاده

جسم گرانجیک

متیاس پذیری

سازگاری پایتون و سیستم آسانش بیان استفاده آنرا، محبوب ترین زبان در جوامع دگرگاه می باشد تبدیل کرده است، مگر از نکات ممکن این هست که زبان را در ساختارهای حقیقی پیشنهاد می کند، متن مدلیند حقیقت را از دست می بینند

تجلیل پذیری و در آسانی علم دارد به دنبال آن مفهوم API می باشد که سازمان های بزرگ در متخصصان علم دارد به دنبال آن مفهوم API می باشد که سازمان

API های آن به صوره پیچیده می باشد آن پایتون را عنوان اول این کاراکترها می دانند

است که دانشمندان و ترسیم دهنگان یا دیگری ماشین را برای ساخت برنامه های ابزارهای کامپیوتری

تجزیه و تحلیل احسات و NL API هایی دارند

Pandas و Numpy - B

Numpy کتابخانه اساسی برای محاسبات دیاضی و علمی است. این کتابخانه آرایه های ماتریس های بزرگ و چند بعدی به صوره میتوانی برای این روش سطح بالا کار بر روی این آرایه های پشتیبانی می کند.

Pandas یکی از کتابخانه های علم دارد است که برای ایجاد ساختار داده استفاده می شود. Pandas از عطاف پذیری قدرتمندی دارد که ایجاد ساختار داده های میتواند این روش را ایجاد کند.

ایجاد می کند و هر کاره می تواند ساختار داده های چند بعدی یا جردی که ناممکن دغیره را ایجاد کند. علاوه بر این، از این کتابخانه برای دست ورزی و تجزیه و تحلیل داده های استفاده می شود.

- چهار Matplotlib برای تجسم داده ها استفاده می شود.

کتابخانه Matplotlib یک کتابخانه شناخته شده و پر استفاده برای رسم نمودار در پایتون است.

لشامن کار تحقیقی استاده می کردند این نرم افزار بسیار قدرتمند و حرفه ای صراحی شده بود و تقریباً تمام نیازهای یک محقق را پوشاند.

اما پایتون تراست مطلب را شدید دارد بد در دل محتفه امروزی با بازگشته آگرچه مطلب بسیار نرم افزارهای جیز نیست اما امروز می بینیم که خیلی از محتفه پایتون را ترجیح می دهند.

یکی داه های رسم نمودار در پایتون است کتابخانه Matplotlib در پایتون است.

کتابخانه های دیگری نیز برای رسم نمودار در پایتون وجود دارند اما Matplotlib پیشنهاد می شود. با استفاده از Matplotlib شما می توانید انواع نمودار های رسم کنید.

نمودار های دو بعدی سه بعدی نمودار هایی نمودار دیجیتال Scatter Plot و

- D Seaborn چهار برای تجسم داده ها استفاده می شود.

Matplotlib یک کتابخانه پایتون است که برای اس-

سانه شده است و به طور مستمر این برای صادر سازی داده ها (استفاده می شود).

دانشمندان داده با استفاده از این کتابخانه می توانند نقشه های خود را ایجاد کنند.

توابع گزینه های ارائه شده توسط Seaborn برای صدور سازی داده ها بسیار زیاد

است.

E - چکونه ص ترانس کیلے Func در Python تعریف بکنید؟

چهار مرحله برای تعریف یک تابع در پاپtron به شرح زیر است:

1: از لحیه بلیدی Def برای اعلام نام به استفاده بگذرد و پس از آن میتوان نام برای تابع خود انتخاب کنید.

2: پارامترها را به تابع اضافه بگذیرد آنرا داخل پرانتز قرار دهید و خط خود را با ترکاوند دو نقطه (:) ببايان پرسانید.

3: عبارت ماین که تابع باید اجرا کند را اضافه کنید.

اگر تابع باید چیزی در خروجی نمایش دهد، تابع خود را با دستور return خاتمه دهید.  
بروون دستور return تابع شاکیه شی None را برای گردانده

نحوه آن در پاپtron به صورت زیر است:

def function\_name(parameters):

Statement(s)

return expression

F - چرا Python در List comprehension استاده صی شود؟

این ساختار به سادگان می دهد تا بروان نیاز به لغتہ های صلووانی و پیچیده

لیست های جزئی یا سریع است ایجاد کنیم List Comprehension در پاپtron

و ساده را می سریع برای ساخت لیست های جدید است. این روش باعث افزایش خواهیم کرد که تعداد خطوط کدنویسی را کند اما برای کدهای پیچیده بتر است

از همان دلیل مردم (حلقه) استفاده کنیم. این روش فقط مخصوص زبان

پایتون نیت و می توان نموده باس از آن در زبان `C++` و `C#` نام برد.

□ - چگونه می توانیم یک `CSV` file را در `Python` خواند؟

< با استفاده از `CSV Reader` >

در اینجا `CSV` با استفاده از متد `open()` در حالت `'r'` باز می شود که

شی فایل را بازی `ترانس` سپس `خوبی` نابه در متغیر `CSV.Reader(file)` ذخیره می شود. متغیر `CSVFile` در حال حاضر یک خواندن `CSV` با فایل باز شده

در اختیار دارد. این خواندن `CSV` به ما امکان می دهد تا داده های فایل خود را با استفاده از

دستورات ساده پایتون، به راحتی مشاهده کنیم.

< با استفاده از `CSV.DictReader` >

این روش مشابه روش قبلی است، با این فایل `CSV` با استفاده از متد `open()` باز

می شود سپس با استفاده از کلاس `Dict Reader` از صادر `CSV` داده های

موجود در فایل `CSV` به یک دیکشنری تگاشته می شود.

< با استفاده (زیر) `Pandas.Read CSV` >

خواندن یک فایل `CSV` با استفاده از تراجم کتابخانه `Pandas` بسیار ساده

و آسان است. برای خواندن داده های فایل `CSV` از متد `Pandas.Read CSV` کتابخانه `Pandas`

استفاده می شود.

XML و JSON - H  
جی تی او دارند؟

XML و JSON هردو دارای فرمت های خوانا برای انسان جهت انتقال داده از صریق شبکه

و مستعمل از زبان هستند. فرمت مزبور از قابلیت ایجاد خواندن از پشتی در میانها پر

(Decade) پشتیبانی ملشده و ترسیم باز نای سختیکن تا بل خواندن هستنده با این حال مدل JSDL نسبت به XML تکنیک لذتمند جدیدتری دارد و جایگزین برای آن محسوب می‌شود. صریحانه مدل را براسان عوامل زیر با XML متفاوت است:

1- XML نسبت به JSDL از ساختار دستور طولانی‌تری برخوردار است، بنابراین

نوشتن JSON برای برنامه‌نویس به زمان کمتری نیاز دارد.

2- برخلاف XML، فرم JSON متنبیر آرایه استنادیه کند.

3- با استفاده از تابع ساده eval زبان Java Script JSON را خواندیا بازیابی

گرد. این در حالی است که برای خواندن XML باید از Parser آلس ام ال استفاده کرد.

# منابع

١٤٠٣/١٢/٢٠

Machine Learning ۱ بخش

Unsupervised Learning و Supervised Learning - A

Hooshio.com

چه خواست دارید؟

Machine Learning در الگوریتم های Feature Scaling چرا - B

sokan.academy.com/blog/feature-scaling-in-machine-learning  
ضروری است؟

Standardization، Normalization چه تفاوت دارند؟

دادگیری ماشین داده: مبانی، معادله های الگوریتم و ابزارها / تابعی و گردآوری میانه و وزن

برای مقایسه بینی داده Min-Max Normalization چرا - C

استفاده می شود؟

Howsam.org/Normalization-in-ml

Machine Learning چیست در الگوریتم های Regularization - D

blog.faradars.org

Z Score Normalization چیست چرا نمایندگی دارد؟

blog.faradars.org

چه مشکلات را under fitting و overfitting - G

موجودی آفرینش Model-building

Cafe tadrис.com // Quera.org

Train/Test Split چرا کاربرد دارد؟ Cross-validation H

Virgo1.io

چگونه کاری کنند؟ Gradient Descent I

Hamruyesh.com

چرا Deep Learning برای پیچیده ترین مسائل استفاده می شود؟ J

Fa.wikipedia.org

1403/12/20

Python Programming 2 بخش

A - چرا Python زبان برنامه نویس سبب علم داده لست?

Cafe tadrис.com

B - چه تفاوتی نام نهاده Pandas و NumPy باشد؟ مباحث، مفاهیم، الگوریتم، ابزارها، تالیف و گردآوری میالدوزان

C - چرا Matplotlib برای تجسم داده ها استفاده می شود؟

D - چرا برای تجسم داده های پیش فتنه کاربرد دارد؟ Seaborn

E - چگونه می تواند یک Python Func تعریف کلندی.

F - چگونه می توانیم الگوریتم ساده ای را تالیف و گردآوری میالدوزان

F چرا List comprehension در Python استفاده می شود.

Sabz danesh . com

- چگونه می توانید یک CSV file را در پایتون خواند.

ما گزینی مانند داده های میان، مقادیر، الگوریتم ها و ابزارها / تالیف و گردآوری میابویم

ML و JSON چه تفاوتی دارند.

WWW.Tahsiljadeh.com