

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

(تمارین درس مباحث ویژه)

مدرس: مهندس احمدزاده

دانشجویان:

عباس کارگر جهرمی / مهدی فرازنده شهرکی

استاندارد سازی (Standardization) : استاندارد سازی فرایند توزیع داده ها است. استاندارد سازی

با تبدیل داده ها را به صفر و در این رابطه تبدیل می کنند. هر چه قدر مقادیر داده ها به سمت صفر نزدیک شوند

میرسد و در این داده ها به این فرایند می گویند. استاندارد سازی فرایند تبدیل داده ها به فرمت (feature) داده ها

با مقیاس و شکل مختلف می تواند داده ها را به یک مقیاس تبدیل می کند.

نرمال سازی (Normalization) : دیگر فرایند از روش های تغییر مقیاس و استفاده از روشی برای استاندارد سازی

Min-Max است. به این ترتیب داده ها به یک مقیاس داده می شود که در این مقیاس آن ها به یک

بازه [۰، ۱] قرار داده می شود. در کل منظور از نرمال سازی تبدیل داده ها به یک بازه [۰، ۱] است. هر چه داده ها

داده ها به یک مقیاس تبدیل می شود و در این بازه [۰، ۱] قرار می گیرد. این امر باعث می شود داده ها

تفاوت داشته باشند و در این فرایند هر چه داده ها به یک مقیاس تبدیل می شود.

کاربرد است (۱) : در این بازه که داده ها به یک مقیاس تبدیل می شود و در این بازه [۰، ۱] قرار می گیرد و در این

داده ها به یک مقیاس تبدیل می شود و در این بازه [۰، ۱] قرار می گیرد و در این بازه [۰، ۱] قرار می گیرد

کاربرد است (۲) : در این بازه که داده ها به یک مقیاس تبدیل می شود و در این بازه [۰، ۱] قرار می گیرد و در این

بازه از استاندارد سازی است (۳) : در این بازه که داده ها به یک مقیاس تبدیل می شود و در این بازه [۰، ۱] قرار می گیرد و در این

Cross-validation یک روش اعتبار سنجی است که به منظور ارزیابی مدل ما یادگیر ما میسر
استفاده می شود. این روش کمک می کند تا اطمینان حاصل کنیم که مدل تنها به داده های آموزش
فامر نیست و می تواند به درستی روی $train$ set و داده های جدید $test$ set نیز عمل کند.

$train/test$ split: در هنگام استفاده از داده ها معمولاً به دو بخش تقسیم می شوند: (۱) داده های
آموزشی برای آموزش مدل (۲) داده های تست برای ارزیابی مدل. در این روش ممکن است مدل
به دلیل ویژگی های خاص داده های آموزش ($overfit$) به گونه ای که به خوبی تطبیق یابد
بنابر این نتیجه ارزیابی ممکن است شباهت زیادی به تقسیم خاص داده ها داشته و عملکرد مدل روی داده های
واقع در دنیای واقعی کمتر از آن خواهد بود. ممکن است تفاوت داشته باشد.

$Cross-validation$: به این مشکل کمک می کند. در این روش داده ها به چندین دسته تقسیم
می شوند و مدل به نوبت بر روی بخش های مختلف آموزش داده می شود و پس از ارزیابی می شود.
این کار کمک می کند تا مدل از داده های مختلف برای آموزش و ارزیابی استفاده کند.

(۱-۲)

روش بیشینه-کمینه (Min-Max) در مثال سازنده ما به این دلیل استفاده می شود

که داده ما را در یک بعد و در سطح قرار دهد، معمولاً این [اوه] یا [اوه-این] کار را

به دلیل مقولت اهمیت دارد.

(۱) مقیاس همگرا (یکسان) به دیگر مقیاس ها: دیگر مقیاس ها با تغییر داده ها، یکسان را بر می آورند.

(۲) یکسان سازی: وقتی داده ها مقیاس های متفاوتی دارند، الگوریتم نمی تواند به درستی بین

دیگر مقیاس ها تفاوت قائل شود. با نرمال سازی داده ها همه دیگر مقیاس ها به یکسان در می آیند و

الگوریتم به درستی کار می کند.

(۳) اقلترین وقت عمل با استفاده از Min-Max عمل عکس در عمل عکس یا دیگر مقیاس می تواند

میسود باید زیر اصل به داده هایی با مقیاس مشابه دسترس دارد و بنابراین احتمال بهینه تر می شود.

(نکته) سرعت بالاتر از داده آموزشی: الگوریتم می تواند به یکباره داده ها را به مقیاس تبدیل

نماید ممکن است سریعتر همگرا می شود.

به طور کلی این روش به ابعاد یک مقیاس ثابت در داده ها و جلوگیری از تأثیر نامعادله

دیگر مقیاس ها با مقیاس به دیگر استفاده می شود.

گرایش (gradient descent) یک الگوریتم است که برای پیدا کردن پارامترهای مدل به هدف کمینه کردن تابع هزینه استفاده می‌شود.

نقود عملکرد گرایش

① تعریف تابع هزینه ② معادله گرایش ③ به روز رسانی پارامترها ④ تکرار مراحل

انواع مختلف گرایش

① گرایش با (batch gradient descent) ② بقا (stochastic)

③ مینی بچ (mini-batch)

مزایا ① یک الگوریتم ساده و کارآمد برای پیدا کردن پارامترها است ② برای مدل‌های پیچیده و

داده‌های بزرگ به راحتی قابل استفاده است.

معایب ① انتخاب ضریب یادگیری بسیار مهم است و اگر آن را به درستی انتخاب نکنیم، عملکرد ما به شدت کاهش می‌یابد.

② ممکن است به راحتی در یک محلی (local minimum) گیر کند، به خصوص در مدل‌های پیچیده.

Z-Score Normalization یکی روش برای مقایسه بین داده‌ها است که در آن داده‌ها به

گونه‌ای تغییر می‌کند که می‌توانیم آن‌ها را مقایسه کنیم. این کار معمولاً برای

ویژگی‌های داده‌ها در مدل‌های یادگیری ماشین انجام می‌شود تا الگوریتم‌ها بتوانند با داده‌ها

مقایسه‌ای که مقایسه‌ای دارند با دقت بیشتری عمل کنند.

کاربرد مدل آن

۱) مدل یادگیری ماشین: بسیاری از الگوریتم‌های یادگیری ماشین مانند

کریستین خطی یا شبکه‌های عصبی به شدت تحت (KNN) محاسبه Z-score تأثیر می‌گذارد.

داده‌ها هستند. استفاده از Normalization می‌تواند عملکرد مدل را بهبود بخشد.

۲) مقایسه ویژگی‌ها: زمانی که ویژگی‌ها مقایسه‌ای دارند که می‌تواند به مقایسه

تکلیف تبدیل شوند و هیچ ویژگی‌ای از اهمیت بیشتری نسبت به دیگری پیدا نکنند.

۳) شبیه‌سازی و آنالیز آماری: به تغییر و تقلیل داده‌ها می‌تواند به شناسایی نقاط پرت

Z-score یا Z-score می‌کند زیرا تقاطع (3-sigma) یا بیش از ۳- دارند معمولاً به عنوان

نقاط پرت مد نظر گرفته می‌شوند.

به نام خدا

(1-1)

Subject:

Page: ()

یادگیری عمیق (deep learning) به دلیل توانایی حل مسئله به فرم خود در پردازش و

یادگیری و استخراج معنی از داده ها و به دلیل پیچیدگی تر از مدل های سنتز و محاسباتی

(1) توانایی یادگیری و استخراج معنی از داده ها و به دلیل قابلیت به حل مسائل پیچیده و

تشفیر (تجزیه و تحلیل زبان طبیعی) و تشخیص گفتار که دیگر مدل های پیچیده و الگوریتم های

(2) دستگاه های محاسباتی پیچیده و با سرعت بالا و قابلیت حل مسائل یادگیری عمیق

میتوانند به طور مؤثرتری در آن ها حل شوند و توانایی حل مسائل را می توانند انجام دهند.

(3) انواع و نیز در انواع مختلف داده ها Deep learning توانایی کار با انواع مختلف داده ها

از جمله داده های ساختاریافته و غیر ساختاریافته را یافته و دارد.

(4) عملکرد در مسائل پیچیده و به دلیل توانایی پیچیده و مانند فرجه های و تشخیص گفتار و

استاد به تصویر و عملکرد بهتر نسبت به روش های سنتز دارد.

یادگیر نظامت شده (supervised):

لیک روپکړو د یادگیر ماشین است که با استفاده از مجموعه داده‌ها و یک چک‌کننده که در فرآیند آموزش قرار می‌گیرد.

یادگیر بدون نظامت (unsupervised):

یادگیر بدون نظامت از الگوریتم‌هایی برای یافتن الگوها و ساختارها در داده‌ها استفاده می‌کند. این الگوریتم‌ها بدون نیاز به برچسب‌دهی داده‌ها کار می‌کنند.

تفاوت در انواع یادگیر ماشین:

در یادگیر نظامت شده، الگوریتم از مجموعه داده‌ها برای یادگیری یک مدل استفاده می‌کند و نتایج را با داده‌های جدید می‌تواند مقایسه کند. در یادگیر بدون نظامت، الگوریتم از داده‌ها برای یافتن الگوها و ساختارها استفاده می‌کند و نتایج را با داده‌های جدید می‌تواند مقایسه کند. در یادگیر بدون نظامت، الگوریتم از داده‌ها برای یافتن الگوها و ساختارها استفاده می‌کند و نتایج را با داده‌های جدید می‌تواند مقایسه کند.

بهنگام فدا

۱-۴

Subject: Page: ()

regularization به یادگیر ماشین یک روش است که به کاهش اثرات بیش برآیند (overfitting)

داده های آموزش بر روی مدل کمک کند و به واقع بیش برآیند تغییرات مشکلات اساسی به یادگیر ماشین

است که دقت مدل به سمت به داده های آموزش و وابسته می شود و به دلیل داده های جدید عمل کرده نمی داند.

regularization با اضافه کردن یک جرمیه به تابع هدف (objective function) که معمولاً شامل فاصله

آموزش جرمیه به دلیل وابسته شدن بیش برآیند می باشد است. مدل را میسر می کنند که از پیچیدگی زیاد

خودداری کنند و به جای آن به یادگیر الگوهای معمول و قابل تفسیر می شود.

علاوه بر این دو روش لاسر regularization وجود دارند L_1 و L_2 که روش L_1

جرمیه به دلیل وابسته شدن اعمال می شود که به منفرد شدن یک هستند و در نتیجه بسیار از وابسته ها صفر می شوند.

و ما به یک مدل فنی ساده می رسیم. اما در روش L_2 جرمیه به دلیل وابسته شدن اعمال می شود که

بزرگتر از صفر هستند و در نتیجه وابسته های کوچکتر نسبت به L_1 حفظ می شوند.

فرا از بقیان بغیر رشته کمرها استوار و کمرنگ

۱) دیکار لایز بقیاس ها، بقیاس غیر تفصیل می کنند که نوعی و بیشتر با سبک بقیاس دایره مشابه

قرار دارند این فرآیند که به آن زمان ساز و دیار گفته می‌شود به عنوان یک فرآیند اجتماعی تلقی گردد.

(۲) بیود عملکرد الکترولیتها: الکترولیتها نیز به دو دسته کاتیون و آنیون تقسیم میشوند. فاسفید یا ساین (SV) (۵۷۸)

کار مرگش، با عیاض، عبدلیه و سهره به نایب مطلق مرشد.

۴) شلر از نایاب دار عطر: مقایس ماں بسیار مفاد - بین ویت کر ماں توانم منبره شلر

تقریر سے دیکھ کر یا کمر ویزر عدیل (NLU) کے حساب سے ملے ہوئے نتائج پر اس کے شکوکے کا اظہار کیا۔

④ رفع سوگیر میل: بقیاس خبر دادهما تفصیل کی کنند کہ عہد دیگر سر ہا در فرامیند یارید تا ائیر بیدایی

داند. بگو آن ویتس را با قیاس به زنده می توانست به شایع میل تالیف نماید بلکه اندک و سیر اعداد الله.

میرزا عباس خاں، صاحب دیر، خیر خواندہ، پیر دیر، قلعہ مند - حاکم دیر، نائب ملک دیر، دارمند

روش های مقایسه زیر ۱) با مقدار مطلق (AMS) ۲) مقایسه زیر (min-max)

(standardization); $\mu = 1$ & (normalization); $\sigma = 1$

مقاومت (Robust)

"به ناکارضا"

(1-6)

Subject: Page: ()

overfitting: در آموزش ما *data science* یک مدل *overfit* مدل تعریف می شود که دیتایین

زیاد دیتایین کمتر دارد و آموزش دارد، که منجر به تغییر در دیتایین مدل به داده های جدید

می شوند.

underfitting: اگر مدل دیتایین کمتر و دیتایین زیاد باشد که کمتر مدل *under fit* است.

یعنی اینکه مدل به حال افتاده دارد و از آن فرایند دیتایین کند، از یادگیری آن مدل به داده های آموزشی

معمود دارد، منجر به یادگیری و غیر توانسته یاد بگیرد و داده ها را کشف کند.

اهل شکل *overfit* و *under fit*:

وقتی داده ما به وقت دیتایین می کند مدل *overfit* می شود و شما می بینید داده ها کمتر تغییر می کند

مدل *under fit* می شود باید ای بل یادگیری بالایی معهود داشته باشد، خوشبختانه یک راه حل

مناسب برای حل این شکل معهود دارد و آن *Validation* است.