

» بے نام خدا «

درس: مباحث و تشریح

استاد: محمد احمد زارہ

رشتہ: مهندس کامپیوٹر

دانشجو: زمر اکبریں امل - زمر امالی

دانشگاه: ملی مہارت اقد شدہ میٹاب

صفحات

مفهرست

بخش اول

251

A. Supervised Learning و Unsupervised Learning چه تفاوتی دارند؟

3

B. چرا Feature Scaling در الگوریتم های Machine Learning ضروری است؟

4

C. Standardization و Normalization چه تفاوتی دارند؟

5

D. چرا Min-Max Normalization برای مقیاس بندی داده ها استفاده می شود؟

6

E. Z-Score Normalization چیست و چرا استفاده می شود؟

7

F. Regularization در الگوریتم Machine Learning چیست؟

7

G. Overfitting و Underfitting چه مشکلاتی را در Model-building ایجاد می کنند؟

8

H. Cross-Validation چرا، و Train/Test split را به چه دردی می خورد؟

9

I. Gradient Descent چگونه کار می کند؟

10

J. چرا Deep Learning برای Mandegar بهترین انتخاب است؟

بخش اول ← Machine learning →

A supervised and unsupervised learning چه تفاوتی دارند؟ دو نوع اصلی یادگیری ماشین هستند که هر کدام کاربردها و ویژگی‌های خاص خود را دارند.

Supervised learning (یادگیری تحت نظارت) - در این نوع یادگیری مدل با استفاده از داده‌های ورودی و خروجی مشخص آموزش می‌بیند. به عبارت دیگر داده شامل ویژگی‌ها (x) و برچسب (y) هستند و هدف مدل پیش‌بینی خروجی بر اساس ورودی‌ها است.

Unsupervised learning (یادگیری بدون نظارت) - در این نوع یادگیری مدل با داده‌های ورودی بدون برچسب آموزش می‌بیند. هدف اصلی این نوع یادگیری کشف الگوها و ساختارها که بتوان از داده‌ها است.

داده‌ها ...

Supervised learning - نیاز به داده‌های برچسب دارد. به عنوان مثال در تشخیص تصویر، داده شامل تصاویر هستند که به هر تصویر برچسب مشخص (مثلاً گربه یا سگ) اختصاص داده شده است.

Unsupervised learning - داده‌ها بدون برچسب هستند. برای مثال در خوشه‌بندی، داده‌ها تنها شامل ویژگی‌هایی هستند و مدل باید خود به تنهایی گروه‌ها یا خوشه‌ها را پیدا کند.

Subject:

هدف ها ...

Supervised learning ← هدف اصلی این روش پیش فرضی مای جدید بر اساس داده های شناخته شده است این روش می تواند شامل طبقه بندی (Classification) یا رگرسیون (Regression) باشد.

Unsupervised learning ← هدف شش آموختن ساختارها یا روابط در داده ها است. این روش می تواند شامل خوشه بندی (Clustering) یا کاهش ابعاد باشد.

نمونه ...

Supervised Learning ← الگوریتم های مانند درخت تصمیم (Decision Tree) و شبکه عصبی (Neural Network) و رگرسیون خطی (Linear Regression).

Unsupervised learning ← الگوریتم های مانند K-Means و خوشه بندی میزبان و تحلیل مولفه های اصلی (PCA) (Principal Component Analysis).

لیست ...

Supervised learning ← معمولاً نیاز به برچسب و منابع بیشتری برای جمع آوری و آماده سازی داده های برچسب داره.

Unsupervised learning ← به دلیل عدم نیاز به برچسب ها، معمولاً سریع تر و گاهی تر در آماده سازی داده ها است، اما تفسیر نتایج ممکن است پیچیده تر باشد.

B. چرا Feature Scaling در الگوریتم های Machine Learning ضروری است؟

Feature Scaling یا مقیاس و نیزش های در الگوریتم های یادگیری ماشین به دلیل های زیر ضروری است.

1 تاثیر مقیاس بر الگوریتم ها - بسیاری از الگوریتم های یادگیری ماشین مانند کلاسیفیکیشن و رگرسیون ها (Gradient Descent) به فاصله ها و مقادیر ورودی حساس هستند و نیزش ها مقیاس های متفاوتی داشته باشند و نیزش های با مقیاس بزرگ تر می توانند تاثیر بیشتری بر مدل داشته باشند و منجر به تغییر نبرای های نادرست شوند.

2 تعمیم پذیری - در الگوریتم های پیش بر جمع سازی مانند رگرسیون خطی و شیمی های عصبی مقیاس سازی می تواند به همگرایی سریع تر کمک کند و نیزش ها با مقیاس مشابه می توانند فرآیند یادگیری را تسهیل کنند و تعداد تکراری های لازم برای رسیدن به حداقل هار را کاهش دهند.

3 کاهش تاثیر نویز - مقیاس سازی می تواند تاثیر نویز در داده ها را کاهش دهد و نیزش های که مقیاس بزرگتری دارند ممکن است نویز بیشتری را به مدل اضافه کنند.

4 افزایش دقت و کارایی - در نهایت با مقیاس سازی مناسب مدل ها می توانند عملکرد بهتری در پیش بین و تصمیم به داده های جدید داشته باشند.

روش های متداول برای مقیاس سازی

Normalization (نرمال سازی) - به مقیاس $[0, 1]$ یا $[-1, 1]$ تبدیل می کند.

Standardization (استاندارد سازی) - و نیزش ها را به گونه ای مقیاس بندی می کند میانی آن

Normalization و Standardization چه تفاوتی دارند؟

هر دو تکنیک هایی هستند که در پردازش داده ها و یادگیری ماشین برای مقایسه دیتا و ویژگی ها استفاده می شوند اما اهداف و روش های متفاوت دارند.

- تفاوت ...

Standardization (استاندارسازی) - این روش به مقیاس دیتا ها به گونه ای است می باشد که انحراف و انحراف معیار آن یک شود.

Normalization (نورمال سازی) - این روش داده ها را به یک بازه مشخص معمولاً $[0, 1]$ یا $[-1, 1]$ مقیاس دیتا می کنند.

- هدف ...

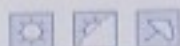
Standardization - هدف اصلی استاندارد سازی ایجاد داده های با توزیع نرمال است که برای برخی از الگوریتم های یادگیری ماشین مانند رگرسیون خطی و SVM بسیار مفید است.

Normalization - هدف نورمال سازی این است که داده ها را در یک محدوده مشخص قرار دهد و برای الگوریتم هایی که مقیاس خاص و حساس هستند (مانند شبکه های عصبی مناسب است).

- استفاده ...

Standardization - معمولاً زمانی استفاده می شود که داده های توزیع نرمالی دارند یا نیاز به حذف فاصله ها و نسبت ها بین داده ها وجود دارد.

Normalization - بیشتر در مواقعی استفاده می شود که داده ها به شدت متنوع هستند یا برای مقایسه داده ها در یک بازه مشخص نیاز به نورمال سازی است.



1. چرا Min-Max Normalization برای مقیاس بندی داده استفاده می شود؟

این تکنیک مقیاس بندی داده ها است که به منظور تغییر مقادیر به یک بازه مشخص بین صفر و یک استفاده می شود.

1 مقیاس یکنواخت ← Min-Max تمام ویژگی ها را به یک مقیاس یکنواخت تبدیل می کند و این باعث می شود مقیاس ویژگی ها بر تاثیر زیادی بر عملکرد مدل داشته باشد.

2 مجبوری عملکرد مدل ← با مقیاس بندی داده ها به یک بازه مشخص توزیع های مانند.

Gradient Descent می تواند سریع تر و با دقت بیشتر همگرا شوند. این به ویژه در توزیع های پارتیالری کمین و شیم های کمین بسیار مهم است.

3 حذف اثر مقیاس ← در داده های دارای ویژگی های با مقیاس های متفاوت، برخی ویژگی ها می توانند بر روی پارامتری مدل تسلط داشته باشند.

4 سادگی و شفافیت ← این روش بسیار ساده است و محاسباتی که به راحتی انجام می شود، برای هر ویژگی فقط کافی است حداقل و حداکثر آن را پیدا کرد و مقادیر نوشت.

5 کاربرد های متنوع ← این روش در زمینه های مختلفی از جمله تحلیل داده ها، یادگیری ماشین، پردازش تصویر، و بسیاری از کاربردهای دیگر استفاده می شود.

6 مناسب برای داده های دارای توزیع یکنواخت ← min-max به خصوص برای داده های که توزیع یکنواخت دارند مناسب است. اگر داده ها به طور طبیعی در یک بازه خاص قرار دارند این روش می تواند به حفظ ساختار داده ها کمک کند.

E Z-Score Normalization چیست و چرا کاربرد دارد؟

استاندارد سازی Z-score یک تکنیک برای مقیاس گذاری داده‌ها است که به منظور مقایسه بیشتر ابعاد و کاهش تاثیر مقادیر پرت (outliers) استفاده می‌شود. این روش به داده‌ها اجازه می‌دهد تا در یک مقیاس استاندارد قرار گیرند که در آن میانگین (mean) داده‌ها برابر با صفر و انحراف معیار برابر با یک است.

کاربرد مهم

1. مقایسه داده‌های مختلف ← این روش کمک می‌کند تا داده‌هایی از توزیع‌ها یا مقیاس‌های مختلف می‌آیند تا به راحتی قابل مقایسه شوند.
2. بهبود عملکرد الگوریتم‌های یادگیری ماشین ← بسیاری از الگوریتم‌ها مانند SVM، K-Nearest به مقیاس داده‌ها حساس هستند این الگوریتم‌ها می‌توانند بهتر عمل کنند.
3. کاهش تاثیر مقادیر پرت ← Z-score به ما کمک می‌کند تا تاثیر مقادیر پرت را بر تحلیل‌های آماری کاهش دهیم.
4. تحلیل‌های آماری → در بسیاری از این تحلیل‌ها فرض بر این است که داده‌های نرمال هستند و Z-score می‌تواند کمک کند تا داده‌ها به این فرض نزدیک‌تر شوند.

F. Regularization در الگوریتم های Machine learning چیست؟

Regularization یا منظم سازی در الگوریتم های یادگیری ماشین Machine learning به مجموعه تکنیک هایی اطلاق می شود که برای جلوگیری از overfitting مدل به کار می روند پیش بهارزش زمانی اتفاق می افتد که مدل یادگیری ماشین به طور خاص به داده های آموزش خود سازگار می شود و توانایی محقق سازی خود را بر روی داده های جدید از دست می دهد.

G. overfitting و Underfitting چه مسئله ای را در Model-building وجود می آورد؟

1. مدل کمترین بر روی داده های جدید - مدل ممکن است در بیشترین دیتا آموزش یافته دقت بالایی داشته باشد اما نتواند داده های جدید کمترین خوبی نداشته باشد.

2. قابلیت تعمیم پایین - مدل نمی تواند به خوبی الگوهای عمومی موجود در داده ها را شناسایی کند زیرا به جزئیات خاص داده های آموزش وابسته است.

3. پیچیدگی غیر ضروری - مدل ممکن است بسیار پیچیده شود که این خود می تواند منجر به افزایش زمان محاسبات و نیاز به منابع بیشتر شود.

Cross Validation چرا در Train/Test split باید داشت؟

اعتبار نسبی همقابل این تست مهم در یادگیری ماشین است که به ما کمک می‌کند تا عملکرد مدل‌های یادگیری ماشین را به طور دقیق‌تری ارزیابی کنیم.

- **دانش واریانس تعیین‌ها...**

در روش‌های ساده Train/Test split تقسیم داده‌ها به دو مجموعه آموزش و آزمون می‌باشد. همچنین به همراه داشته باشند. اگر داده‌ها به طور تصادفی تقسیم شوند ممکن است مجموعه آزمون شامل نمونه‌های غیر معمول باشد که می‌تواند به ارزیابی مدل تاثیر منفی بگذارد.

- **استفاده بهینه از داده...**

همه داده‌ها در نهایت هم در مجموعه آموزش و هم در مجموعه آزمون استفاده می‌شوند. این به ما کمک می‌کند تا از داده‌ها موجود به بهترین نحو استفاده کنیم و در نتیجه مدل‌های بهتری بسازیم.

- **ارزیابی دقیق‌تر مدل...**

با استفاده از Cross Validation ما می‌توانیم عملکرد مدل را در چندین زیرمجموعه از داده‌ها بررسی کنیم و میانگین نتایج را به عنوان ارزیابی نهایی در نظر بگیریم. این کار به ما اجازه می‌دهد تا مدل را در برابر تغییرات مختلف داده‌ها آزمون کنیم و ارزیابی دقیق‌تری داشته باشیم.

- **پیشگیری از Overfitting**

با ارزیابی مدل بر روی داده‌های مختلف Cross Validation می‌توانیم به راحتی متوجه شویم که آیا مدل به خوبی تعمیم می‌یابد یا فقط بر روی داده‌های آموزش به خوبی عمل می‌کند. این به ما کمک می‌کند تا اشتباه‌های بهتری در مورد تعمیمات مدل و اشتباه و بیزنس ما داشته باشیم.

Gradient Descent چگونه کار میکند؟

تئردیانتی کاهش Gradient Descent یک روش بهینه سازی است که به طور گسترده در یادگیری ماشین و بهینه سازی توابع استفاده می شود. هدف این روش یافتن مقادیر بهینه ضرایب یا پارامترهاست که منجر به کمترین خطای ممکن در این مدل ها می شود. در اینجا مراحل کار در تئردیانتی کاهش را به شما می گویم...

تصرف توابع هزینه عم تابع است که میزان آن حقله یا هزینه پیش بین مدل را ارزیابی می کند
هدف ما این است که این تابع را به حداقل برسانیم.

محاسبات برای آنکه هر دایره تابع هر چند است که نشان می دهد شیب تابع در نقاط مختلف است و محاسبات برای آنکه به حاصل می رسد که در کدام محبت باید حدیث بود تا هر چند را با هم در هم.

به وزیرستان پیرامتر هام با استفاده از بردیانت محاسب شده پیرامترهای بیروزیستان می شوند

تقریر فرمایند. این مدارحل به علوه فصول انجام می شود تا زمانی که تابع هزینه به حداقل

ممكن بعد يا تغييرات تابع هترینیم بم مقدار مشفق باعش یا بد. این روند عقلی است و فایده
باشد و به اشتراک مناسب تفرق یادگیری بیشتر دارد.

انواع تدریس ۱۰۰

۱- تم ادبیات بافتی / مامل ... در این نوشتن تم ادبیات بر اساس تقاسم داده ها و آمارهای موجود است.

۲. تمرینات بافتش (تقویت بافت) • بر اساس این قانون تقویت (فرض از داده واقعی) میسر شود و می تواند سرعت

۲۰. آیات یافتن صفت مهم: تمکین از دوروش بال که به اساس زیر صیغهای از داده و صیغهای شود.

حل ۱: چرا Deep learning برای پیچیده ترین مسائل استفاده می شود؟

به دلیل ویژگی ها و قابلیت های خاص که دارد برای حل پیچیده ترین مسائل در حوزه های مختلف استفاده می شود.

توانایی یادگیری در داده های بزرگ

Deep learning به طور خاص برای کار با مجموعه های داده بزرگ بهینه سازی شده است. با افزایش حجم داده ها عملکرد این الگوریتم ها بهبود می یابد.

قابلیت تعمیم

مدل های یادگیری عمیق می تواند به خوبی تعمیم یابند و در شرایط جدید عملکرد خوبی داشته باشند. این ویژگی به ویژه در مسائل پیچیده و داده های آموزش محدود است. تنوع زیاد باشد.

حل ۲: های پیشرفت

مدل های پیشرفته ای مانند Transformer در پردازش زبان طبیعی و GANs در تولید داده های جدید مثال دهنده پیشرفت های اخیر هستند که به حل مسائل پیچیده کمک می کنند. موارد کاربرد گسترده

Deep learning در بسیاری از زمینه ها مانند بینایی کامپیوتری، پردازش زبان طبیعی، تشخیص صدا و حتی بازی ویدیویی موفقیت های چشمگیری داشته باشد. این تنوع کاربرد نشان دهنده قدرت و انعطاف این تکنیک است.