

Date: / /

Sat. Sun. Mon. Tue. Thu. Wed. Fri.

Subject: -----

درس: مباحث ویژه

موضوع: machine learning

نام دانشجو:

ترنس کولی پور

کبری رحیمی زاده

نام استاد: محمد احمد زاده

ترم بهمن ۱۴۰۳

Date: / /

Sat. Sun. Mon. Tue. Thu. Wed. Fri.

Subject: -----

## Machine Learning.

### A - Supervised Learning و Unsupervised Learning تفاوتی دارند؟

تفاوتی اصلی بین Supervised Learning و Unsupervised Learning در نوع داده‌هایی است که برای آموزش

مدل استفاده می‌شوند.

#### 1. Supervised Learning (یادگیری تحت نظارت):

- در این روش، داده‌ها شامل ورودی‌ها و خروجی‌های مشخص هستند. شما برای هر ورودی جواب درست را

دارید. هدف این است که مدل یاد بگیرد چگونه از ورودی‌ها به خروجی‌ها برسد.

#### 2. Unsupervised Learning (یادگیری بدون نظارت):

- در این روش، شما فقط داده‌های ورودی را دارید و هیچ خروجی مشخص وجود ندارد.

- هدف این است که مدل خودش الگوها و سافت‌های موجود در داده‌ها را شناسایی کند.



Date: / /

Sat Sun Mon Tue Wed Fri

Subject: -----

## B- Feature scaling در الگوریتم های Learning ضروری است:

بیا مقیاس بندی ویژگی های از مایل هم در Feature Scaling بیشتر بدازش داده ها در الگوریتم های

یادگیری ماشین است. این کار به چند دلیل ضروری است:

1. تسویه فرآیند یادگیری: وقتی که ویژگی ها در مقیاس های مختلف باشند، برضا از الگوریتم ها مثل

Gradient Descent ممکن است به آرامی همگرا شوند. بیا مقیاس بندی، سرعت یادگیری افزایش می یابد.

2. بهبود عملکرد مدل: برای الگوریتم ها مثل K-Nearest Neighbors (KNN) و

Support Vector Machines (SVM)، به فاصله ها وابسته اند. اگر ویژگی ها مقیاس های متفاوتی

داشته باشند، ممکن است مدل به ویژگی های در مقیاس بزرگتری وابسته تر شود و در نتیجه دقت کمتری

داشته باشد.

3. جلوگیری از سوگیری: اگر یکی از ویژگی ها مقیاس بزرگی داشته باشد، می تواند بر نتایج نهایی تاثیر

بگذارد و باعث سوگیری در مدل شود. بیا مقیاس بندی، تمام ویژگی ها به یک سطح نزدیک می شوند.

Date: / /

Sat Sun Mon Tue Wed Fri

Subject: -----

.C

هر دو تکنیک های Standardization و Normalization بین بردارش داده ها هستند که برای مقیاس

بندی ویژگی ها در یادگیری ماشین استفاده می شوند، اما تفاوت هایی دارند.

### 1. Standardization (استاندارد سازی):

در این روش، داده ها به گونه ای مقیاس بندی می شوند که میانگین آن ها صفر و انحراف معیار آن ها یک

باشد. یعنی به فرمول زیر عمل می کنیم:

$$Z = \frac{(X - \mu)}{\sigma}$$

]

این روش بیشتر برای داده هایی که توزیع نرمال (Gaussian) دارند مناسب است.

### 2. Normalization (نرمال سازی):

در این روش، داده ها به مقیاس متناهی (یا کمتر) تبدیل می شوند، معمولاً از فرمول زیر استفاده می شود:

$$X' = \frac{(X - \text{min}(X))}{(\text{max}(X) - \text{min}(X))}$$

]

این روش برای داده هایی که دارای مقادیر متفاوتی هستند مفید خواهد بود و ویژگی های بیرونی دیگر را کم کند و مناسب تر است.



Date: / /

Sat Sun Mon Tue Thu Wed Fri

Subject: -----

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

برای مقیاس بندی داده ها به دلایل زیر Min-Max Normalization استفاده می شود:

1. یکسان سازی مقیاس: این روش کمک می کند که تمام ویژگی ها در یک بازه مشخص، معمولاً بین 0 تا 1 قرار بگیرند.

2. افزایش دقت مدل: با نرمال سازی داده ها، می توانیم عملکرد مدل های یادگیری ماشین را بهبود

بخشیم. الگوریتم هایی مانند K-Means یا K-Nearest Neighbors به مقیاس داده ها حساس

هستند و نرمال سازی می تواند دقت آن ها را افزایش دهد.

3. سرعت همگرایی: در یادگیری عمیق، نرمال سازی داده ها می تواند به تسریع روند یادگیری کمک کند و زمان آموزش را کاهش دهد.

4. حفظ نسبت ها: با نرمال سازی، نسبت های بین داده ها حفظ می شود. این مهمی که سافت و توزیع داده ها تغییر نمی کنند.

5. جلوگیری از اثرات ناخواسته: اگر داده ها در مقیاس های مختلف باشند، ممکن است بعضی ویژگی ها تأثیر بیشتری بر

نتیجه نهایی بگذارند. Min-Max Normalization برای مشکل راحل می کند.

چنانچه بخواهیم موثر و ساده برای Min-Max Normalization مقیاس بندی داده ها است که در بسیاری از پروژه های

یادگیری ماشین و تحلیل داده ها کاربرد دارد.

Date: / /

Sat. Sun. Mon. Tue. Thu. Wed. Fri.

Subject: -----

## E. Z-Score Normalization چیست و چرا کاربردش دارد؟

یکی از روش‌های نرمال سازی Z-Score Normalization داده‌هاست که به کمک آن می‌توان

مقادیر کمی مجموعه داده را به یک مقیاس مشابه تبدیل کرد. این روش به ویژه در تحلیل داده‌ها مدل سازی آماری کاربرد دارد.

کاربردها:

1. مقایسه داده‌ها: با نرمال سازی داده‌ها می‌توان مقادیر مختلف را به راحتی مقایسه کرد، حتی مقیاس‌های

متفاوتی داشته باشند.

2. بهبود عملکرد مدل‌ها: در الگوریتم‌های یادگیری ماشین نرمال سازی داده‌ها می‌تواند به بهبود دقت و سرعت

آموزش مدل کمک کند.

3. شناسایی نقاط نااهنجار: با استفاده از Z-Score می‌توان نقاط دور افتاده یا نااهنجار را شناسایی

کرد. زیرا Z-score این نقاط معمولاً بسیار بزرگ یا کوچک ظاهر می‌شود.

به طور کلی، Z-Score Normalization ابزاری قدرتمند برای تحلیل داده‌ها و بهبود کیفیت مدل سازی است.



Date: / /

Sat Sun Mon Tue Wed Fri

Subject: -----

Regularization در الگوریتم‌های machine learning چیست؟

Regularization در الگوریتم‌های یادگیری ماشین به تکنیک‌هایی اشاره دارد که برای جلوگیری از

Overfitting استفاده می‌شوند و وقتی که یک مدل خیلی خوب بر روی داده‌های آموزشی عمل می‌کند اما در داده‌هایی

جدید یا تست عملکرد خوبی ندارد، به این وضعیت Overfitting می‌گویند.

Regularization به روش‌هایی گفته می‌شود که سادگی مدل را افزایش می‌دهد و از پیچیدگی آن کم می‌کند. دو نوع اصلی

Regularization عبارتند از:

1. L1 Regularization (Lasso): این روش باعث کردن مجموع مقادیر مطلق پارامترها به تابع هزینه، بدین ترتیب از

پارامترها را به صفر رسانده و این کار باعث می‌شود که مدل ساده‌تر و قابل تفسیرتر شود.

2. L2 Regularization (Ridge): در این روش، مجموع مربعات پارامترها به تابع هزینه اضافه

می‌شود. این باعث می‌شود که Overfitting پارامترها به طور یکنواخت کوچک شود و از جلوگیری کمتر.



Date: / /

Sat Sun Mon Tue Thu Wed Fri

Subject: -----

6. Overfitting و Underfitting چه مشکلاتی را در model-building

با وجود یادگیری دارند Underfitting و overfitting دو مشکل عمده در مدل سازی

یادگیری ماشین هستند که می توانند تا غیر از منفی بر عملکرد مدل داشته باشند.

overfitting: این مشکل زمانی اتفاق می افتد که مدل به قدری پیچیده باشد که نتواند

الگوهای واقعی را یاد بگیرد، بلکه به نویز و داده های غیر ضروری نیز حواسش شود.

علائم overfitting: 1. دقت بالا در داده های آموزش و دقت پایین در

داده های تست.

2. پیچیدگی بیش از حد مدل (مثل تعداد زیاد پارامترها).

Underfitting: این مشکل زمانی رخ می دهد که مدل به قدری ساده باشد که نتواند

الگوهای موجود در داده ها را به درستی یاد بگیرد. در این حالت مدل حتی در داده های

آموزشی نیز عملکرد ضعیفی دارد. و معمولاً به دلیل انتخاب نادرست مدل یا پارامترها

ناکافی ایجاد می شود.



Date: / /

Sat. Sun. Mon. Tue. Thu. Wed. Fri.

Subject: -----

4.  $cross-validation$  چرا در  $train/test split$  کاربرد دارد؟

$cross-validation$  (اعبار سنجی متقاطع) یکی از تکنیک‌هاست که در یادگیری ماشین استفاده می‌شود.

که به ما کمک می‌کند تا مدل‌ها را بهتر ارزیابی کنیم. وقتی از  $train/test split$  استفاده می‌کنیم، داده‌ها را به دو بخش تقسیم می‌کنیم: یکی برای آموزش مدل و یکی برای تست اما

ممکن است این تقسیم‌بندی به صورت تصادفی باعث شود که نتایج ما تحت تأثیر قرار بگیرد.

1.  $Gradient Descent$  چگونه کار می‌کند؟

$Gradient Descent$  (تیرا دیان گرادیان) یک الگوریتم بهینه‌سازی است که برای پیدا کردن حداقل

یک تابع استفاده می‌شود. این الگوریتم معمولاً در یادگیری ماشین و شبکه‌های عصبی به کار می‌رود.

1. انتخاب یک نقطه شروع: اولین قدم، انتخاب یک نقطه تصادفی در فضای پارامترهاست.

2. محاسبه گرادیان: گرادیان تابع هزینه را محاسبه می‌کنیم. گرادیان نشان‌دهنده شیب تابع

است و نشان می‌دهد که در کدام جهت باید حرکت کنیم تا هزینه کاهش یابد.

Date: / /

Sat. Sun. Mon. Tue. Thu. Wed. Fri.

Subject: -----

۳ به روز رسانی پارامترها: با استفاده از گرادینت، پارامترها را به روز رسانی می‌کنیم. ~~فصل~~

۴ تکرار: این مراحل را تکرار می‌کنیم تا زمانی که به یک حداقل محلی برسیم یا تغییرات پارامترها

به اندازه کافی کوچک شود.

۵ پایان: وقتی که به حداقل تابع هزینه رسیدیم یا تعداد تکرارها به حد اکثر رسید،

الگوریتم متوقف می‌شود.

۶ چرا Deep Learning برای پیچیده‌ترین مسائل استفاده می‌شود؟

Deep Learning به خاطر توانایی آن در یادگیری الگوهای پیچیده و استخراج ویژگی‌ها از داده‌ها

از داده‌های بزرگ و متنوع، برای پیچیده‌ترین مسائل استفاده می‌شود. این روش به

شکل‌های عصبی عمیق متکی است که می‌توانند لایه‌های متعددی از پردازش را انجام دهند

و این باعث می‌شود که بتوانند اطلاعات را به شکل‌های مختلف تجزیه و تحلیل کنند.