

ect:

نام خدا

Date:

دانشگاه ملی مهارت  
آفریننده میناب

موضوع: (machine learning) تک

نام و نام خانوادگی: زهرا دروغوار، سریم رفیعی زاده، سریم سجادی

واحد درسی: مباحث ویژه

رشته: مهندسی کامپیوتر

مدرس: آقای محمد احمدزاده

۱۴۰۲

بهار

۱. supervised learning و unsupervised learning چه تفاوتی دارند؟

در یادگیری با نظارت (supervised learning)، مدل با استفاده از داده‌های دارای برچسب آموزش

می‌بیند تا خروجی را پیش‌بینی کند. در یادگیری بدون نظارت (unsupervised learning) مدل، الگوها را در

داده‌های بدون برچسب کشف می‌کند.

۲. چرا Feature Scaling در الگوریتم‌های machine learning ضروری است؟

Feature Scaling در یادگیری ماشین ضروری است زیرا:

۱. مقیاس‌های متفاوت ویژگی‌ها ممکن است مقیاس‌های متفاوتی داشته باشند که باعث شود الگوریتم

به ویژگی‌هایی با مقادیر بزرگتر اهمیت بیشتری بدهد.

۲. عملکرد بهتر: بسیاری از الگوریتم‌ها (مانند SVM، رگرسیون لجستیک، شبکه‌های عصبی) به مقیاس

ویژگی‌ها حساس هستند و Scaling می‌تواند باعث بهبود سرعت و دقت آن‌ها شود.



۳ Standardization و Normalization چه تفاوتی دارند؟

• Standardization: میانگین را صفر و انحراف معیار را یک می‌کند (توزیع نرمال).

• Normalization: مقادیر را به بازه‌ای بین ۰ و ۱ می‌برد.

• انتخاب: بستگی به داده و الگوریتم دارد.

Standardization برای داده‌های نرمال و Normalization برای داده‌های با توزیع نامنظم و مقیاس‌های

با مقادیر خیلی بزرگ مناسب‌تر است.

۴ چرا Min-Max Normalization برای مقایسه بندی داده‌ها استفاده می‌شود؟

Min-Max Normalization برای مقایسه بندی داده‌ها استفاده می‌شود زیرا:

۱. محدود کردن دامنه: مقادیر را به بازه‌ای مشخص (معمولاً ۰ تا ۱) محدود می‌کند که در بسیاری از الگوریتم‌ها، مانند شبکه‌های عصبی، به بهبود عملکرد کمک می‌کند.

۲. حفظ روابط: روابط بین داده‌ها حفظ می‌شود و فاصله‌ها نسبت به یکدیگر متناسب‌تر می‌مانند.

۳. داده‌های متنوع: برای داده‌هایی با مقادیر مختلف و مقیاس‌های متفاوت مناسب است و باعث جلوگیری از غلبه یک یا چند ویژگی بر ویژگی‌های دیگر می‌شود.

Date:

Subject:

۴. محافظت از نمرات جدول: تغییرات شدید در دامنه مقایسه را کاهش می دهد و نمرات جدول توزیع را در برنجی الگوریتم ها تسهیل می کند.

۵. Z-score Normalization چیست؟ و چرا کاربرد دارد؟

Z-score Normalization رویی برای مقایسه بندی داده ها است که هر مقدار را با استفاده از میانگین (μ) و انحراف (σ) ویژگی نرمال می کند.

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

کاربردها:

1. توزیع نرمال: برای داده هایی که به توزیع نرمال نزدیک اند مناسب است.

2. مقایسه ویژگی ها: امکان مقایسه ویژگی های با مقیاس های مختلف را فراهم می آورد.

3. شناسایی ناهنجاری ها: کمک به شناسایی داده های دور افتاده (outlier).

4. بهبود عملکرد الگوریتم ها: به بهبود رقت در الگوریتم های یادگیری ماشین کمک می کند.



## ۴ Regularization در الگوریتم‌های machine learning چیست؟

Regularization در یادگیری ماشین به تکنیک‌هایی گفته می‌شود که برای کاهش overfitting و بهبود

تعمیم پذیری مدل‌ها به کار می‌روند. این کار با افزودن عبارات به تابع هزینه انجام می‌شود.

Regularization ابزاری کلیدی برای بهبود کیفیت مدل‌های یادگیری ماشین است.

## ۵ overfitting و underfitting چه معنایی را در Model-building بوجود می‌آورند؟

overfitting و underfitting دو مشکل رایج در ساختن مدل‌های یادگیری ماشین هستند که می‌توانند عملکرد

مدل را به طرز چشمگیری تحت تأثیر قرار دهند.

### overfitting:

• تعریف: مدل بیش از حد به داده‌های آموزشی برازش می‌شود و الگوهای تصادفی و نویز را یاد می‌گیرد.

• مشکلات:

• عملکرد ضعیف روی داده‌های جدید یا تست.

• فقدان تعمیم پذیری، یعنی مدل نمی‌تواند به خوبی تعمیم پیدا کند.

### underfitting:

• تعریف: مدل قادر به یادگیری الگوهای موجود در داده‌های آموزشی نیست و خیلی ساده است.

• مشکلات:

• دقت پایین در هر دو داده‌های آموزش و تست.

• عدم توانایی مدل در شناسایی روابط معنادار.

1- cross-validation چرا در Train / Test split کاربرد دارد؟

cross-validation (اعتبارسنجی) (مقایسه) برای تخمین دقیق تر عملکرد مدل و جلوگیری از overfitting در

مقایسه با یک Train / Test split ساده استفاده می شود.

Train / Test split فقط یکبار داده ها را تقسیم می کند، در حالی که cross-validation چندین بار این کار را تکرار می کند و

میانگین نتایج را محاسبه می کند تا از ناپایداری قابل اعتمادتری بدست آید.

4- Gradient Descent چگونه کار می کند؟

گزاره این کاهش دهنده است. الگوریتم بهینه سازی برای یافتن کمترین مقدار یک تابع (معمولاً تابع هزینه در یادگیری ماشین).

محاسبه گرایی شیب تابع در هر نقطه و حرکت در خلاف جهت شیب (با گام های کوچک) به پایین ترین نقطه توسط نرخ یادگیری).

در اکثر موارد به روز رسانی می کند تا به تدریج به کمینه برسد.



ject:

Date:

۱۰. چرا Deep learning برای پیچیده ترین مسائل استفاده می شود؟

دلایل استفاده از یادگیری عمیق Deep learning برای مسائل پیچیده:

• استخراج خودکار ویژگی ها: خودکشی ویژگی های مهم را از داده می گیرد، نیازی به مهندسی دستی نیست.

• مقیاس پذیری: با داده: هرچی داده بیشتر باشد، عملکردش بهتر می شود.

• توانایی مدل سازی پیچیدگی: با شبکه ها عصبی عمیق، الگوهای پیچیده را در داده ها می تواند یاد بگیرد.