

«ب نام خدا»

Machine Learning : ۱ بخش

نام درس : مباحث ویژه

Python Programming : ۲ بخش

سرآئین :

نام دانشجو : معصومه فریج

واژه سرم : ورودی بهمن ۱۴۰۳

نام مدرس : استاد محمد احمدزاده

۱۳۰۳، ۱۲، ۲۰

## بخش ۱: machine Learning

A. Supervised Learning و Unsupervised Learning چه تفاوتی دارند؟

Supervised Learning (یادگیری تحت نظارت) برای یادگیری از داده‌های برچسب‌خورده استفاده دارد. جایی که مدل با استفاده از ورودی‌ها و خروجی‌ها را مشخص می‌کند. Unsupervised Learning (یادگیری بدون نظارت) برای داده‌های بدون برچسب مربوط به داده‌ها استفاده می‌کند. جایی که مدل باید الگوها و ساختارها را بیابد. اینها را می‌توان به دو دسته تقسیم کرد: Classification (طبقه‌بندی) و Regression (رگرسیون).

B. چرا Feature Scaling در الگوریتم machine Learning ضروری است؟

Feature Scaling (مقیاس‌بندی ویژگی‌ها) برای اطمینان از اینکه مقیاس ویژگی‌ها در مدل machine Learning مشابه قرار دارند، ضروری است. این کار به الگوریتم‌ها کمک می‌کند تا به سرعت و به درستی یاد بگیرند و از تأثیر ویژگی‌ها با دامنه‌های مختلف جلوگیری می‌کند.

C. Normalization و Standardization چه تفاوتی دارند؟

Standardization: داده‌ها را با استفاده از میانگین و انحراف معیار مقیاس‌بندی می‌کند. این فرآیند داده‌ها را با میانگین صفر و انحراف معیار یک متغیر می‌کند.

Normalization: داده‌ها را به یک دامنه مشخص (مثلاً  $[0, 1]$ ) مقیاس‌بندی می‌کند.

این کار معمولاً برای داده‌هایی که توزیع نرمال ندارند، مفید است.

D. چرا Min-max Normalization برای مقیاس‌بندی داده‌ها استفاده می‌شود؟

Min-max Normalization: داده‌ها را به یک دامنه مشخص بین حداقل و حداکثر مقیاس‌بندی می‌کند. این روش برای کمالات تأثیر مقادیر دور از اقل و اکثر (outliers) و برای الگوریتم‌هایی که به مقیاس وابسته‌اند، مفید است.

۱. کتابخانه sklearn (مثلاً sklearn.preprocessing) برای مقیاس‌بندی داده‌ها استفاده می‌کند.



## E. Z-score Normalization چیست و چرا کاربرد دارد؟

تعمی مقیاس چند است که با استفاده از میانگین و انحراف معیار Z-score Normalization معروف است. داده‌ها را به مقیاس استاندارد تبدیل می‌کند. این روش به شما می‌دهد مقادیر دور از انتظارات کم می‌کند و در الگوریتم‌هایی که فرض نرمال بودن داده‌ها دارند، کاربرد دارد.

## F. Regularization در الگوریتم‌های Machine Learning چیست؟

بیشتر برای جلوگیری از مدل‌های Overfitting تکنیکی است که برای جلوگیری از Overfitting استفاده می‌شود. با افزودن یک جریمه به تابع هزینه مدل را تشویق می‌کند که پارامترهای داشته باشد و از پیچیدگی بیش از حد جلوگیری کند.

## G. Overfitting و Underfitting چه مشکلاتی را در Model-building

و وجود می‌آورند؟

Overfitting: مدل بیش از حد به داده‌های آموزشی تطبیق پیدا می‌کند و نتواند به جدید تعمیم یابد.

Underfitting: مدل به اندازه کافی با داده‌ها تطبیق پیدا نمی‌کند و نتواند الگوهای موجود در داده‌ها شناسایی کند.

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

Cross-validation به ارزیابی و تعیین سازش مدل‌ها کمک می‌کند و با تعمیم داده‌های چندین بخش از داده‌های آموزشی و تست به صورت جعبه استفاده می‌کند. این کار به شما کمک می‌کند تا از تعمیم پذیری مدل‌ها اطمینان حاصل کنید.

۱. Gradient Descent چگونه کار میکند؟

یکی از الگوریتم‌های بهینه‌سازی است که با محاسبه گرادیان تابع هزینه نسبت به پارامترها، به سمت کمینه تابع حرکت می‌کند. فرآیند به تدریج پارامترها را به روز می‌کند تا بهترین مدل ممکن را پیدا کند.

۲. چرا Deep Learning برای پیچیده‌ترین مسائل استفاده می‌شود؟  
Deep Learning به دلیل توانایی در استخراج ویژگی‌های پیچیده و الگوهای غیرخطی در داده‌ها، برای مسائل پیچیده مانند شناسایی تصویر، پردازش زبان طبیعی و یادگیری تقویتی استفاده می‌شود.

شبکه‌های عصبی عمیق می‌توانند به صورت خودکار ویژگی‌های مهم را استخراج کنند و به همین دلیل در پردازش داده‌های جزئی و پیچیده بسیار موثر هستند.



۱۴۰۳ / ۱۲ / ۲۵

## بخش ۲ Python Programming

A. چرا Python زبان برنامه نویسی محبوب علم است؟

Python به دلیل سادگی و خوانایی کد، کتابخانه‌های قوی و متنوع (مانند Pandas

NumPy، Matplotlib، SciPy)

SciPy) و جامعه بزرگ و فعال، یکی از محبوب‌ترین زبان‌ها برای علم

داده تبدیل شده است. همچنین Python قابلیت‌های خوبی برای تجزیه و تحلیل

داده‌ها، یادگیری ماشین و علم داده دارد و به راحتی با سایر زبان‌ها و ابزارها ادغام می‌شود.

B. NumPy و Pandas چه تفاوتی دارند؟

NumPy: یک کتابخانه برای محاسبات عددی است که به ویژه برای آرایه‌ها و ماتریس‌ها

عملیات ریاضی و جبر خطی دارد.

Pandas: یک کتابخانه برای تجزیه و تحلیل و دستکاری داده‌ها است که به طور

خاص برای کار با داده‌های جدولی (DataFrame) مناسب است و امکاناتی برای تجزیه

و تحلیل داده‌ها، زمان‌بندی و کار با داده‌های ناقص ارائه می‌دهد.

C. چرا Matplotlib برای تجزیه داده‌ها استفاده می‌شود؟

یک کتابخانه قدرتمند برای تجزیه داده‌ها است که امکان ایجاد انواع مختلف Matplotlib

۱۱ نمودارها و گراف‌ها را فراهم می‌کند. این کتابخانه به ویژه برای تولید نمودارهای

دو بعدی و سه بعدی با دیگر ابزارها مانند Jupyter Notebook و RStudio

سفارشی‌سازی کرد. همچنین ادغام می‌شود و امکان تجزیه داده‌ها به طور

مؤثر را با NumPy و Pandas کتابخانه‌ها مانند Jupyter Notebook فراهم می‌کند.

D. Seaborn چرا برای تبصر داده‌های پیچیده تر استفاده می‌شود؟

ساخت شده و امکانات Matplotlib یک کتابخانه تبصر داده‌ها است که بر پایه

Seaborn پیچیده تر برای تبصر و تحلیل و تبصر داده‌ها ارائه می‌دهند. این

کتابخانه به ویژه برای کار با داده‌های پیچیده و چند بعدی مناسب است و به کاربران اجازه

می‌دهد تا نمودارهای زیبا تر و واضح تر (DuZuFum) ایجاد کنند. اما، با Seaborn

Pandas پیچیده تر به نظر می‌رسد.

E. چگونه می‌توانید یک Function در Python تعریف کنید؟

با کلمه کلیدی def تعریف می‌شود. نکته تائید یلوئی تاکید است که متغیرهای تعریف می‌شود

F. چرا List Comprehension در Python استفاده می‌شود؟

List Comprehension یک روش مختصر و آسان برای ایجاد لیست‌ها در Python

است. این روش به شما اجازه می‌دهد تا به سادگی و به سرعت لیست‌ها را ایجاد کنید.

نقشه‌های برای ایجاد لیست است.

G. چگونه می‌توانید یک CSV file را در Python خواند؟

برای خواندن یک فایل CSV می‌توان از کتابخانه Pandas استفاده کرد. به عنوان مثال

```
import pandas as pd
```

```
data = pd.read_csv
```

```
('file.csv')
```



۱.۱ JSON و XML چه تفاوتی دارند؟

سادگی: JSON ساده تر و خوانا تر از XML است. JSON به طور طبیعی به زبان های برنامه نویسی مدرن نزدیک تر است.

معمول: JSON معمولاً متبصر کمتر نسبت به XML دارد، زیرا دارای نشانه گذاری کمتر است.

نوع داده: JSON از انواع داده های مختلف (مانند رشته ها، اعداد، آرایه ها و اشیاء) پشتیبانی می کند، در حالی که XML بیشتر بر اساس نشانه گذاری است و نوع داده ها به طور ضمنی تعریف می شود.