

« به نام حقدا »

موضوع درس : مباحث ویدیه

استاد: محمد احمدزاده

دانشجو: محمدرضا خیرج

بخش ۱ | machine Learning :

Supervised Learning (یادگیری تحت نظارت) - یادگیری از داده‌های برچسب

معمومینند. در مقابل، **unsupervised learning** (یادگیری بدون نظارت) به کار

با دادن هار بدون جریمه میوه ها می شود. عیبی که مدل باید الگوها و ساختار هار می نمایند، و در داده ها شناسایی کنند.

8. چرا Feature scaling، نورملائزیشن machine Learning فلوں است؟

مقیاس بندر و بزرگساز (برای اطمینان از اینکه همه ویژگی‌ها در یک Feature scale مقیاس مشابه قرار دارند، ضروری است. این کار با استاندارد کردن همه مقیاس‌ها و سریع تر یاد بگیرند و از تاثیر و بزرگسازها کمتر جلوگیری می‌کند).

• Normalization, Standardization

Standardization: داده‌ها را با استفاده از میانگین و انحراف معیار مقیاس بند می‌کنند. این فرآیند داده‌ها را با میانگین صفر و انحراف معیار یک استاندارد می‌کند.

Normalization: داده‌ها را به یک دامنه مشخص (مثلاً $[0, 1]$) مقیاس بندی می‌کند. این کار معمولاً برای داده‌هایی که توزیع نرمال ندارند، مفید است.

د. جبار Min-max Normalization برای مقیاس بندی داده ها استفاده می شود.

Min-max Normalization: داده‌ها را به یک دامنه مشخص بین حداقل و حداکثر مقادیر تبدیل می‌کند.

میکنند. این روش براساس تأثیر مقادیر دور از انتقال (outliers) و براساس الگوریتم

هاین که به مقیاس وابسته اند. ۱- پایان در مثل $K_{N,N}$ چهار مفید است

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

تعمی مقیاس چند است که با استفاده از میانگین و انحراف معیار Z-score Normalization

oh مرورگرش داده ها را به مقیاس استاندارد تبدیل میکند. این روش به شناسایی مقادیر

دور از انتظار کمک میکند و در الگوریتم هایی که فرض نرمال بودن داده ها دارند، کاربرد دارد.

F. Regularization در الگوریتم های machine Learning چیست؟

بیشتر برای جلوگیری از مدل ها و overfitting تکنیکی است که برای جلوگیری از

overfitting استفاده می شود. 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 می آورند؟

overfitting: مدل بیش از حد به داده های آموزشی تطبیق پیدا میکند و نتواند به

جدید تعمیم یابد.

underfitting: مدل به اندازه کافی با داده ها تطبیق پیدا نمی کند و نتواند الگوهای

موجود در داده ها شناسایی کند.

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

cross-validation به ارزیابی و تعیین سازش مدل ها کمک میکند و با تقسیم داده های

چندین بخش، از داده های آموزشی و تست به صورت جعبه استفاده میکند. این کار

به کاهش واریانس و اطمینان از تعمیم پذیری مدل کمک میکند.

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 شناسایی 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؟
 داده

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

NumPy, Matplotlib, SciPy

SciPy) و جامعه بزرگ و فعال، یکی از محبوب‌ترین زبان‌های برنامه نویسی 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 کار با آرایه‌ها و ماتریس‌ها

بسیار کارآمد است. این کتابخانه عملگرهای ریاضی و عملیات آرایه‌ای را به سرعت انجام می‌دهد.

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

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

دو بعدی و سه بعدی با دیگر کتابخانه‌ها مانند NumPy و Pandas به راحتی ادغام می‌شود.

سخت‌ترین سازش کرد. همچنین ادغام می‌شود و امکان تجزیه داده‌ها به طور

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

D. Seaborn چرا برای تجسم دادن های پیچیده تر از برادرش Matplotlib ساخته شده و امکانات Matplotlib یک کتابخانه تجسم داده ها است که بر پایه Seaborn پیچیده تر برای سبک و تقایل و تجسم دادن هارا که میدهند. این کتابخانه به ویژه برای کار با داده های پیچیده و چند بعدی مناسب است و به کاربران اجازه میده تا نمودارهای زیباتر و راحت تر را با Dazufuam ایجاد کنند و با Seaborn Pandas پیچیده تر بهارند.

E. چگونه می توانید یک Function در Python تعریف کنید؟ برای این، از اعلان کنید با کلمه کلیدی def تعریف می شود. نکته ثانیه اینکه این آرایه است که متغیرهای اجرایی شود.

F. چرا List Comprehension در Python استفاده می شود؟ List Comprehension یک روش مختصر و آسان برای ایجاد لیست ها در Python است. این روش که اعلانها تر و کوتاه تر میکند و معمولاً سریع تر استفاده لازم ایجاد می کند. تعریف برای ایجاد لیست است.

G. چگونه می توانید یک CSV file را در Python خواند؟ برای خواندن یک فایل CSV می توان از کتابخانه Pandas استفاده کرد. به عنوان مثال:

```
import pandas as pd
data = pd.read_csv('file.csv')
```


۱. H. JSON و XML به 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.

عجم: JSON معمولاً عجم تر نسبت به XML دارد، زیرا دارای فاصله گذاری کمتری است.

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