

تمرین سوم

دیتاساینس - دکتر نادری

:Q1

در این کد، شما با استفاده از مدل‌های درخت تصمیم (Decision Tree) و جنگل تصادفی (Random Forest)، قصد دارید پیش‌بینی کنید که آیا یک مشتری در صورت دریافت وام، آن را پرداخت خواهد کرد یا خیر. داده‌ها از یک مؤسسه مالی گرفته شده‌اند و شامل ویژگی‌هایی همچون نرخ بهره، درآمد سالانه، نسبت بدهی به درآمد، و امتیاز اعتباری FICO هستند. سوالات مختلفی در این کد مطرح شده است، از جمله بررسی و تحلیل داده‌ها، آموزش مدل‌ها، پیش‌بینی و ارزیابی مدل‌ها با استفاده از معیارهای مختلفی مانند ROC-AUC، F1-Score، و Classification Report.

هدف این است که شما با انجام این مراحل، توانایی مدل‌ها را در پیش‌بینی بازپرداخت وام ارزیابی کنید.

:Q2

می‌خواهیم با استفاده از داده‌های مربوط به ساختمان‌های شهر نیویورک، ENERGY STAR Score را برای آن‌ها پیش‌بینی کنیم.

الف) از آنجایی که EDA بخش ثابت همه پروژه‌های Machine Learning است، در این بخش ابتدا شما با استفاده مطالب گذشته موارد خواسته شده داخل نوت بوک Machine_Learning_Q_Part1 را کامل کنید. (15)

ب) حال با استفاده از نتایج بخش قبل مدل خود را داخل قسمت دوم (Machine_Learning_Q_Part2) پیاده سازی کنید. (25)

Q3:

در این تمرین، شما باید رگرسیون خطی منظم‌شده را پیاده‌سازی کنید و از آن برای مطالعه مدل‌هایی با ویژگی‌های مختلف بایاس-واریانس استفاده کنید. در ابتدا داده‌ها را از فایل `ex5data1.mat` بارگذاری کرده و آن‌ها را بصورت گرافیکی نمایش دهید. سپس، با استفاده از تابع `linear_reg_cost_function()` هزینه و گرادیان برای رگرسیون خطی منظم‌شده محاسبه کنید. پس از آن، مدل رگرسیون خطی را بدون منظم‌سازی آموزش داده و منحنی یادگیری را برای مدل رگرسیون خطی و رگرسیون چندجمله‌ای ترسیم کنید. در نهایت، با استفاده از منحنی اعتبارسنجی، بهترین مقدار پارامتر منظم‌سازی (λ) را انتخاب کنید.

موفق باشید.