تمرین 12 / 4500 امتیاز / Logistic Reg

۱ . روی دیتاست زیر مرحله به مرحله گام های زیر را اجرا کنید. (دیتاست در ریپوی مدرسه موجود است!)

breast–cancer.csv

- **بخش ۱ — آماده‌سازی محیط**
- تمام کتابخانه‌های موردنیاز را برای کار با داده، آموزش مدل و ارزیابی وارد کن.

- **بخش ۲ — بارگذاری و مشاهده داده‌ها**
- دیتاست breast_cancer را از ماژول sklearn.datasets بارگذاری کن.
- داده را به یک DataFrame پانداس تبدیل کن تا بتوانی راحت‌تر آن را مشاهده و تحلیل کنی.
- پنج سطر اول دیتاست را چاپ کن.
- بررسی کن که ستون هدف (target) چه مقادیری دارد و هر مقدار نشان‌دهنده‌ی چه نوع توموری است (benign یا malignant).
- تعداد ردیف‌ها و ستون‌های دیتاست را بررسی کن.

مدرسه هوش مصنوعی فیلاگر

- **بخش 3 – بررسی مقدماتی داده‌ها**
- توزیع کلاس‌ها (تعداد نمونه‌های سرطانی و غیرسرطانی) را محاسبه کن.
- اطلاعات کلی دیتاست (نوع داده، nullها، dtypeها) را با متد info() مشاهده کن.
- بررسی کن که آیا در داده‌ها مقدار گمشده وجود دارد یا خیر.
- در صورت وجود مقدار گمشده، راه‌حلی برای حذف یا جایگزینی آن ارائه کن.

- **بخش 4 – آماده‌سازی داده برای مدل**
- ویژگی‌ها (X) را از ستون هدف (y) جدا کن.
- داده‌ها را به دو بخش آموزش و تست تقسیم کن (۸۰٪ آموزش، ۲۰٪ تست).
- اندازه‌ی هر بخش را چاپ کن تا از صحت تقسیم مطمئن شوی.

مدرسه هوش مصنوعی فیلاگر

- **بخش 5 – ساخت و آموزش مدل**
- یک شیء از کلاس LogisticRegression ایجاد کن.
- مدل را روی داده‌های آموزش fit کن.
- در صورت بروز خطای همگرایی، پارامتر max_iter را افزایش بده.
- پس از آموزش، ضرایب مدل را چاپ و تفسیر کن.

- **بخش 6 – پیش‌بینی و ارزیابی**
- خروجی پیش‌بینی مدل را برای داده‌های تست محاسبه کن.
- دقت مدل را با استفاده از accuracy_score محاسبه کن.
- ماتریس درهم‌ریختگی (confusion matrix) را چاپ و تحلیل کن.
- گزارش طبقه‌بندی (classification_report) را چاپ کن و مقادیر precision و recall را مقایسه کن.
- توضیح بده در مسائل پزشکی کدام شاخص مهم‌تر است و چرا.

مدرسه هوش مصنوعی فیلاگر

- **بخش 7 – بررسی احتمال‌ها**
- با استفاده از predict_proba احتمال سرطانی بودن یا نبودن را برای چند نمونه چاپ کن.
- یکی از نمونه‌ها را انتخاب کن و پیش‌بینی مدل را با مقدار واقعی مقایسه کن.

- **بخش 8 – بهبود مدل**
- داده‌ها را با استفاده از StandardScaler نرمال‌سازی کن.
- مدل را مجدداً آموزش بده و دقت آن را قبل و بعد از نرمال‌سازی مقایسه کن.

- **بخش 9 – ذخیره و بارگذاری مدل**
- مدل آموزش‌دیده را با استفاده از joblib در فایل cancer_logistic_model.pkl ذخیره کن.
- در یک اسکریپت جداگانه، مدل ذخیره‌شده را مجدداً بارگذاری کن.
- با استفاده از ورودی جدید (مثلاً یکی از ردیف‌های تست)، پیش‌بینی جدید انجام بده

پایان

مدرسه هوش مصنوعی فیلاگر