تمرین ۱: آیا Accuracy همیشه معیار مناسبی است؟

سناریو:

یک سرویس ایمیل، مدلی برای **تشخیص ایمیل‌های اسپم (Spam)** طراحی کرده است.

این مدل روی **1000 ایمیل** آزمایش شده و نتیجه زیر به دست آمده است.

Confusion Matrix:

	Predicted Spam	Predicted Normal
Actual Spam	35	15
Actual Normal	20	930

بخش اول — Implement

الف) با استفاده از اطلاعات بالا، موارد زیر را محاسبه کنید.

- TP
- TN
- FP
- FN

ب) Accuracy را محاسبه کنید.

ج) Precision را محاسبه کنید.

د) Recall را محاسبه کنید.

هـ) F1-Score را محاسبه کنید.

بخش دوم — Observe

نتایج به‌دست‌آمده را در جدول زیر وارد کنید.

Metric	Value
Accuracy	
Precision	
Recall	
F1-score	

بخش سوم — Reason

الف) با وجود اینکه Accuracy مدل بسیار بالا است، آیا می‌توان گفت این مدل برای تشخیص ایمیل‌های اسپم مناسب است؟ دلیل خود را توضیح دهید.

ب) اگر هزینه از دست دادن یک ایمیل اسپم (قرار گرفتن آن در Inbox) زیاد باشد، کدام معیار نسبت به Accuracy اهمیت بیشتری پیدا می‌کند؟

ج) اگر هدف این باشد که **ایمیل‌های عادی کاربران به اشتباه اسپم تشخیص داده نشوند**، کدام معیار اهمیت بیشتری پیدا می‌کند؟

د) در این مسئله، کدام نوع خطا (False Positive یا False Negative) پرهزینه‌تر است؟ چرا؟

تمرین ۲: ارزیابی کامل یک مدل طبقه‌بندی

از دیتاست **Breast Cancer Wisconsin** موجود در کتابخانه sklearn.datasets استفاده کنید.

```
from sklearn.datasets import load_breast_cancer

data = load_breast_cancer()
```

بخش اول — Implement

الف) داده‌ها را به نسبت 80٪ آموزش و 20٪ آزمون تقسیم کنید.

(از random_state=42 استفاده کنید.)

ب) یک مدل Logistic Regression آموزش دهید.

ج) روی داده آزمون، مقادیر زیر را محاسبه کنید.

- Accuracy
- Precision
- Recall
- F1-score

د) Confusion Matrix را رسم کنید.

هـ) گزارش کامل classification_report را نمایش دهید.

بخش دوم — Observe

الف) نتایج معیارها را در جدول زیر وارد کنید.

Metric	Value
Accuracy	
Precision	
Recall	
F1-score	

ب) از روی Confusion Matrix مشخص کنید:

- چند نمونه False Positive وجود دارد؟
- چند نمونه False Negative وجود دارد؟

ج) کلاس‌بندی مدل برای کدام کلاس عملکرد بهتری داشته است؟

پاسخ خود را بر اساس classification_report توضیح دهید.

بخش سوم — Threshold Experiment

خروجی احتمالات مدل را با استفاده از دستور زیر استخراج کنید.

```
y_prob = model.predict_proba(X_test)[:, 1]
```

سپس برای Thresholdهای زیر، پیش‌بینی مدل را محاسبه کنید.

0.3
0.5
0.7

برای هر Threshold، مقادیر زیر را محاسبه کنید.

- Precision
- Recall
- F1-score

.نتایج را در جدول زیر ثبت کنید

Threshold	Precision	Recall	F1-score
0.3			
0.5			
0.7			

بخش چهارم — Reason

الف) با افزایش Threshold، Precision چه تغییری می‌کند؟ آیا این رفتار با انتظار شما سازگار است؟

ب) با افزایش Threshold، Recall چه تغیری می‌کند؟

ج) اگر این مدل برای تشخیص سرطان استفاده شود، کدام Threshold را ترجیح می‌دهید؟

پاسخ خود را با استناد به نتایج به‌دست‌آمده توضیح دهید.

د) اگر همین مدل برای فیلتر کردن ایمیل‌های اسپم استفاده شود، آیا همان Threshold مناسب است؟

دلیل خود را بیان کنید.

تمرین ۳: آیا مدل شما قدرت تعمیم (Generalization) دارد؟

از داده‌های مصنوعی زیر استفاده کنید.

```
from sklearn.datasets import make_moons

X, y = make_moons(
    n_samples=500,
    noise=0.25,
    random_state=42
)
```

داده‌ها را به نسبت **80٪ آموزش** و **20٪ آزمون** تقسیم کنید.

بخش اول — Implement

سه مدل زیر را آموزش دهید.

مدل اول

```
DecisionTreeClassifier(max_depth=1)
```

مدل دوم

```
DecisionTreeClassifier(max_depth=5)
```

مدل سوم

```
DecisionTreeClassifier() (بدون محدودیت عمق)
```

برای هر مدل موارد زیر را محاسبه کنید.

- روی داده آموزش Accuracy
- روی داده آزمون Accuracy

بخش دوم — Observe

نتایج را در جدول زیر وارد کنید.

Model	Train Accuracy	Test Accuracy
Depth = 1		
Depth = 5		
Full Tree		

در ادامه، مرز تصمیم (Decision Boundary) هر سه مدل را رسم کنید.

بخش سوم — Reason

الف) کدام مدل دچار Underfitting شده است؟

بر اساس نتایج جدول و نمودارها توضیح دهید.

ب) کدام مدل دچار Overfitting شده است؟

علت را توضیح دهید.

ج) کدام مدل بهترین Generalization را دارد؟

پاسخ خود را بر اساس اختلاف عملکرد Train و Test بیان کنید.

د) چرا داشتن Accuracy برابر با 100٪ روی داده آموزش لزوماً به معنی عملکرد خوب مدل نیست؟

هـ) فرض کنید بخواهید از این مدل در یک سامانه واقعی استفاده کنید. کدام مدل را انتخاب می‌کنید؟ چرا؟

تمرین ۴: بررسی Bias–Variance Tradeoff

از دیتاست **Breast Cancer Wisconsin** استفاده کنید.

```
from sklearn.datasets import load_breast_cancer

data = load_breast_cancer()
```

داده‌ها را به سه بخش تقسیم کنید.

- Train : 60%
- Validation : 20%
- Test : 20%

از random_state=42 استفاده کنید.

بخش اول — Implement

مدل‌های زیر را آموزش دهید:

```
DecisionTreeClassifier(max_depth=1)

DecisionTreeClassifier(max_depth=2)

DecisionTreeClassifier(max_depth=3)

DecisionTreeClassifier(max_depth=5)

DecisionTreeClassifier(max_depth=10)

DecisionTreeClassifier()
```

برای هر مدل موارد زیر را محاسبه کنید.

- Train Accuracy
- Validation Accuracy

بخش دوم — Observe

نتایج را در جدول زیر وارد کنید.

Max Depth	Train Accuracy	Validation Accuracy
1		
2		
3		
5		
10		
None		

سپس نموداری رسم کنید که:

- محور افقی: max_depth
- محور عمودی: Accuracy

و هر دو منحنی Train و Validation را روی یک نمودار نمایش دهید.

بخش سوم — Reason

الف) با افزایش عمق درخت، Accuracy روی داده آموزش چه تغییری می‌کند؟

ب) Accuracy روی داده Validation چه رفتاری دارد؟ آیا همیشه افزایش پیدا می‌کند؟

ج) کدام مقدار max_depth بهترین انتخاب است؟ پاسخ خود را بر اساس نمودار توضیح دهید.

د) اگر مدلی دارای Accuracy بسیار زیاد روی Train و Accuracy پایین روی Validation باشد، این وضعیت نشان‌دهنده چه مفهومی است؟

هـ) کدام مدل دارای Variance بیشتری است؟ استدلال خود را بر اساس نتایج ارائه دهید.

و) اکنون **بهترین مدل را بر اساس Validation انتخاب کنید** و فقط یک بار آن را روی مجموعه **Test** ارزیابی کنید.

Accuracy به‌دست‌آمده را گزارش کرده و توضیح دهید که چرا از مجموعه Test تنها در مرحله نهایی استفاده کردید.