

作业5：机器学习-分类算法-逻辑回归

要求：

- 在Python Jupyter中回答，需运行出结果。
- 网页-打印-另存为pdf，命名：作业5-班级-学号后两位-姓名.pdf
- 按时上传职教云，过期不补。

任务一：逻辑回归可视化：是否录取模型（分类）

- 数据：data/ex2data1.txt
- 字段1：exam1，科目1成绩
- 字段2：exam2，科目2成绩
- 字段3：Admitted，是否被录取，1被录取，0未录取。

```
In [19]: # 导入必要的包
```

1. 数据加载

```
In [20]: # 数据加载
```

```
Out[20]:
```

	Exam 1	Exam 2	Admitted
0	34.623660	78.024693	0
1	30.286711	43.894998	0
2	35.847409	72.902198	0
3	60.182599	86.308552	1
4	79.032736	75.344376	1

```
In [21]: # 查看数据基本信息
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   Exam 1      100 non-null   float64
1   Exam 2      100 non-null   float64
2   Admitted    100 non-null   int64  
dtypes: float64(2), int64(1)
memory usage: 2.5 KB
```

```
In [22]: # 查看目标字段的数据分布
```

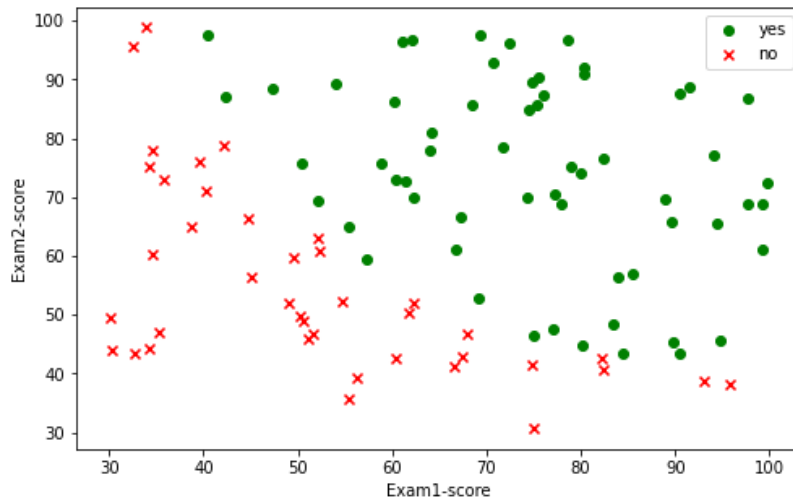
```
Out[22]:
```

1	60
0	40

Name: Admitted, dtype: int64

```
In [23]: # 数据可视化: 通过的用绿色o表示, 未通过用红色x表示, 观察数据
```

```
Out[23]: Text(0, 0.5, 'Exam2-score')
```



2. 数据集拆分

```
In [28]: # 生成特征矩阵X, 目标变量y
```

```
In [29]: # 拆分训练集和测试集, 分别为0.8, 0.2, 随机种子666
```

```
Out[29]: ((80, 2), (20, 2), (80,), (20,))
```

3. 数据处理 (这里数据两个字段, 范围接近, 可以不做标准化)

4. 构建模型

```
In [31]: # 构建逻辑回归模型
# 用训练数据 训练模型
# 输出模型参数
print('逻辑回归的参数w和b分别是: ', )
```

```
逻辑回归的参数w和b分别是: [[0.22756937 0.19779402]] [-25.85537181]
```

5. 模型评估

```
In [33]: # 模型评估
# 混淆矩阵
# 分类报告: 包含了精度, 召回率和f1值

# 在训练集上的效果

print('数据集上的混淆矩阵: \n',)
print('\n数据集上的分类报告: \n',)
```

数据集上的混淆矩阵:

```
[[34  6]
 [ 3 57]]
```

数据集上的分类报告:

	precision	recall	f1-score	support
0	0.92	0.85	0.88	40
1	0.90	0.95	0.93	60
accuracy			0.91	100
macro avg	0.91	0.90	0.90	100
weighted avg	0.91	0.91	0.91	100

```
In [34]: # 在测试集上的效果

print('训练集上的混淆矩阵: \n',)
print('\n训练集上的分类报告: \n',)
```

训练集上的混淆矩阵:

```
[[ 7  2]
 [ 0 11]]
```

训练集上的分类报告:

	precision	recall	f1-score	support
0	1.00	0.78	0.88	9
1	0.85	1.00	0.92	11
accuracy			0.90	20
macro avg	0.92	0.89	0.90	20
weighted avg	0.92	0.90	0.90	20

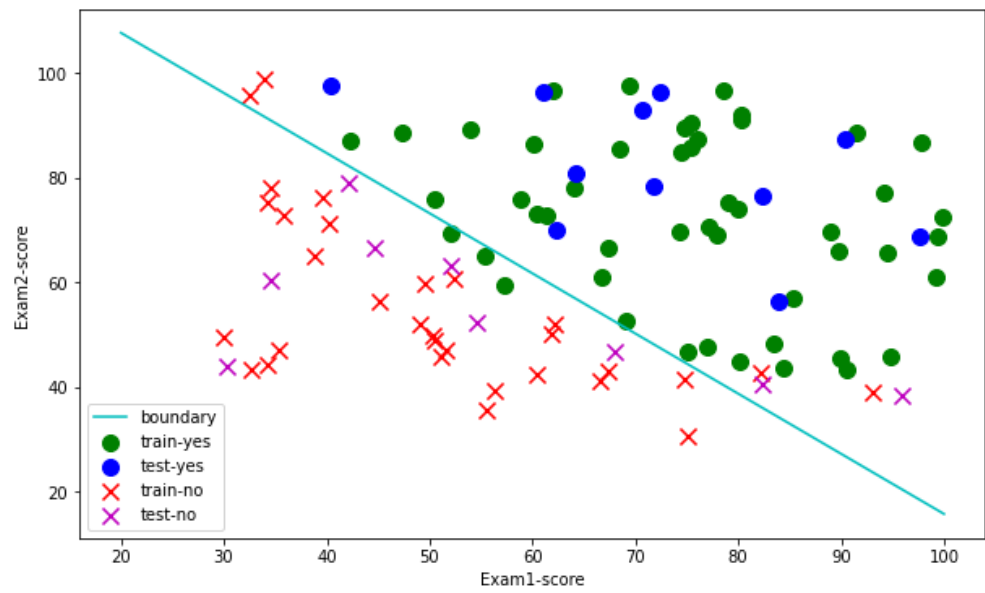
6. 模型可视化 (选做)

画出模型的边界线, 以及训练集和测试集的数据。边界线: $w_1x_1 + w_2x_2 + b = 0$, $x_2 = -(w_1x_1 + b) / w_2$

对照上面的训练集和测试集的分类报告, 找到对应关系。

```
In [38]: # 画出模型的分界线, 以及训练集和测试集的数据
```

```
Out[38]: Text(0, 0.5, 'Exam2-score')
```



```
In [ ]:
```

```
In [ ]:
```