

中小学智慧作业研究实践案例

一、基本信息			
申报人	沈毓骏	工作单位	上海电子信息职业技术学院
授课年级	中职2 年级		
二、实践案例			
案例题目	你能在泰坦尼克海难中生存吗？ - 回归神经网络的交互式体验		
学科名称	人工智能入门	应用的平台系统	Python + Jupyter
案例简介 (200 字 以内)	<p>本案例利用 Jupyter 工具搭建了一个泰坦尼克海难幸存者推理系统，根据所有乘客的各项数据构建乘客的幸存概率模型。本案例作为人工智能入门课程中机器学习启蒙单元，学生在 jupyter 平台上可以完全参与数据采集、数据可视化、数据分析、模型构建、模型训练、模型推理等各个环节的互动，进行数据选择，参数修改等交互操作。从而使学生深刻了解机器学习神经网络的工作原理，并且通过该案例懂得尊重生命的基本道德要求。</p>		
实践应用 介绍	<p>包括但不限于案例的目的、内容、实施方法、实践中遇到的问题……</p> <p>(此部分为重点阐述内容)</p> <p>案例的目的：</p> <p>本案例通过让学生全程体验机器学习神经网络的搭建、训练、使用，使学生能够具体的理解比较抽象的神经网络结构和机器学习过程，使学生更好的接受人工智能的启蒙普及，为之后的学术发展或职业发展打下基础。</p> <p>案例的内容：</p> <p>本案例首先利用 Jupyter 平台设计了一个基于浏览器的一体化教学载体，在案例中采用浏览器页面显示技术图文并茂的向学生展示了泰坦尼克号海难事故的背景、逻辑回归的概念描述、乘客的可视化数据分析、幸存回归网络的构建、幸存回归网络的训练及自定义乘客幸存推演。页面中内嵌实现代码（代码采用 Python 开发，机器学习采用 TensorFlow 实现）。</p>		

项目一 生存预测

任务三 泰坦尼克生存预测

1.3.1 任务背景

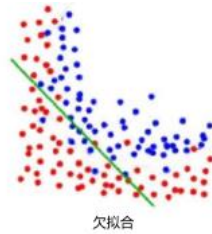
相信大家听说过泰坦尼克号游轮遇难事件。泰坦尼克号是当时世界上体积最庞大、内部设施最豪华的客运轮船，有“永不沉没”的美誉。然而不幸的是，在它的处女航中，泰坦尼克号便遭厄运——它从英国南安普敦出发驶向美国纽约。1912年4月14日23时40分左右，泰坦尼克号与一座冰山相撞，造成右舷船舱至船中部破裂，五间水密舱进水。4月15日凌晨2时20分左右，泰坦尼克船体断裂成两截后沉入大西洋底3700米处。2224名船员及乘客中，1517人丧生，其中仅333具罹难者遗体被寻回。



数据科学家收集了1309名船员及乘客的个人信息和存活情况。我们可以通过这些个人数据信息来分析和最后存活情况的关联性。在形成模型后我们甚至可以预测虚拟人物的存活情况，比如电影中的Jack和Rose，看看他们的预测结果和电影中是否一致。



图 1: 事故背景介绍



我们以复习考试为例：

- 欠拟合：光看书不做题觉得自己会了，上了考场啥都不会。
- 过拟合：做课后题全都能做对，上了考场还是啥都不会。
- 一般的拟合：做了题，背了老师画了重点，考试60分过了。
- 优秀的拟合：课后题全能作对，考试100分。

我们以训练一个树叶种类模型为例：

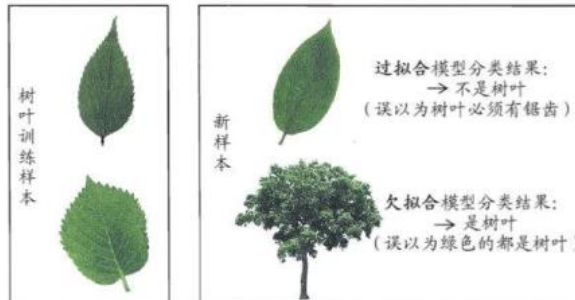


图 2: 模型拟合问题描述

In [5]:

```
survived_df = df_data[df_data['survived'] == 1]
survived_df[['pclass', 'survived']].groupby(['pclass']).sum()
plt.figure(figsize=(10, 5))
plt.subplot(121)
sns.countplot(x='pclass', data=survived_df)
plt.title('Pclass Survived')
plt.ylabel('Survived Count')
plt.subplot(122)
plt.pie(survived_df[['pclass', 'survived']].groupby(['pclass']).sum(),
        labels=['1', '2', '3'], autopct='%1.0f%%')
plt.show()
```

<ipython-input-5-01ec5349cef6>:9: MatplotlibDeprecationWarning: Non-ID inputs to pie() are currently squeezed(), but this behavior is deprecated since 3.1 and will be removed in 3.3; pass a 1D array instead.

```
plt.pie(survived_df[['pclass', 'survived']].groupby(['pclass']).sum(),
```

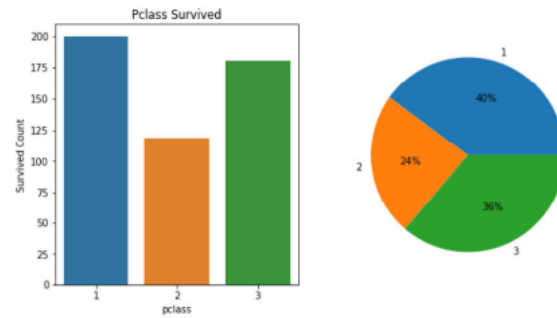


图 3:数据的可视化分析

In [20]:

```
from ipywidgets import *
pclass = widgets.Dropdown(options=[1, 2, 3], description='船舱等级', value=2)
sex = widgets.RadioButtons(options=['female', 'male'], description='性别', value='female')
age = widgets.IntSlider(min=0, max=80, step=5, description='年龄', value=30)
sibsp = widgets.Dropdown(options=[0, 1, 2, 3], description='船上兄弟姐妹', value=0)
parch = widgets.Dropdown(options=[0, 1, 2, 3], description='船上父母子女', value=0)
fare = widgets.Dropdown(options=[5, 20, 50, 100, 200, 300], description='票价', value=50)
embarked = widgets.Dropdown(options=['C', 'Q', 'S'], description='登船港口', value='C')
btn = widgets.Button(description='预测')
display(pclass, sex, age, sibsp, parch, fare, embarked, btn)

def dis(obj):
    person_info = [1, pclass.value, sex.value, age.value, sibsp.value, parch.value, fare.value, embarked.value]
    new_passenger_pd = pd.DataFrame([person_info], columns=selected_cols)
    all_passage_pd = selected_df_data.append(new_passenger_pd)
    x_features, y_label = prepare_data(all_passage_pd)
    surv_probability = model.predict(x_features)
    all_passage_pd.insert(len(all_passage_pd.columns), 'surv_probability', surv_probability)
    print('自定义乘客的幸存概率为: ')
    print(surv_probability[-1:])

btn.on_click(dis)
```

船舱等级

性别 female
 male

年龄

船上兄弟姐妹

船上父母子女

票价

登船港口

预测

自定义乘客的幸存概率为:
[[0.9857078]]

图 4: 乘客幸存推演

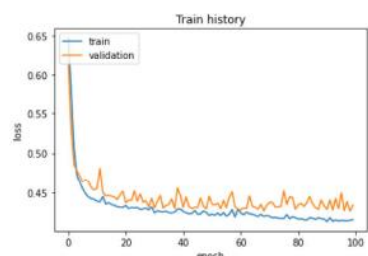
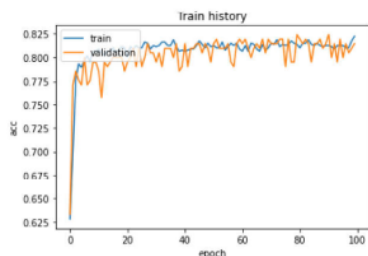
所设计的学生作业包括：

1. 分类模型过拟合和欠拟合的头脑风暴讨论以及汇报。
2. 乘客数据可视化分析中数据项筛选代码修改和数据项必要性的讨论
3. 幸存回归模型的训练参数修改以及参数对推理结果影响的分析。

案例的实施方法：

教学实施（4 课时）：

1. 以电影《泰坦尼克号》作为课程导入介绍泰坦尼克号海难事故的情况。
2. 教师和学生都以 Jupyter 为载体代替传统的纸质教材和 PPT 进行知识点讲解。
3. 讲解模型过拟合和欠拟合概念时以 jupyter 上学校考试为类比引导学生进行头脑风暴讨论并由小组代表进行汇报发言。
4. 讲解乘客数据可视化分析时引导学生对数据项的必要性进行讨论，如：乘客船票的价位是否影响其幸存率，乘客的性别是否影响其幸存率等。学生可直接在 jupyter 上修改代码中数据项并重新运行和查看可视化分析结果。
5. 讲解模型训练部分的时候结合前章的教学内容引导学生对模型代码中优化器选择、训练次数等参数进行修改，通过自动生成的模型训练曲线图对训练结果进行比较从而选择合适的优化器和次数参数。



6. 模型训练完毕后指导学生以电影中角色 Jack 和 Rose 的数据进行幸存推理验证模型的有效性。

然后我们把电影中Jack的个人信息组成一组数据送入模型，观察其生存的概率。通过电影情节我们获知Jack 23岁，在小酒馆赢了了一张三等舱的船票从南安普顿上船。电影中的结局是长眠于大西洋底。运行下面的代码观察Jack的生存概率。



In [18]:

```
Jack_info = [0, 3, 'male', 23, 0, 0, 5.000, 'S']
new_passenger_pd = pd.DataFrame([Jack_info], columns=selected_cols)
all_passage_pd = selected_df_data.append(new_passenger_pd)
x_features, y_label = prepare_data(all_passage_pd)
surv_probability = model.predict(x_features)
all_passage_pd.insert(len(all_passage_pd.columns), 'surv_probability', surv_probability)
all_passage_pd[-1:]
```

Out [18]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	surv_probability
0	0	3	1	23.0	0	0	5.0	2	0.136822

图 5: Jack 的验证结果

三、专家意见

推荐意见

专家签字

年 月 日