**2021年安徽省大数据与人工智能应用竞赛**

**人工智能赛道-网络赛试题（本科组）**

**第一部分：人工智能基础环境搭建部署**（15分）

注：任务1与任务2任选一题完成即可。

**🞏任务1**：Anaconda 3、scikit-learn、OpenCV 3.X、PyTorch 1.8.X、torchvision 0.9.X库的安装与配置。

🖉要求（1）需使用比赛平台提供的上述各软件安装包进行安装与配置。

（2）在配置完成的环境中运行task1\_1.py，将输出结果截图放入答题卡对应位置。

**🞏任务2**：Anaconda 3、scikit-learn、OpenCV 3.X、TensorFlow 2.x库的安装与配置。

🖉要求（1）需使用比赛平台提供的上述各软件安装包进行安装与配置。

（2）在配置完成的环境中运行task1\_2.py，将输出结果截图放入答题卡对应位置。

**第二部分：样本数据预处理**（30分）

**🞏任务**：按照要求对给定csv格式数据进行处理。

task2.csv文件为本题的数据文件。利用Python编写代码完成以下任务：

（1）分别求取Max TemperatureF和Min TemperatureF两个特征的均值和方差。

（2）对Mean TemperatureF特征进行标准化，输出**标准化之后**的Mean TemperatureF特征的**均值**。

标准化采用计算公式如下：

其中，表示标准化之前的特征，表示标准化之后的特征。

🖉说明：将代码源文件与运行结果截图放入答题卡对应位置。

**第三部分：传统机器学习算法设计及应用**（20分）

**🞏任务：**根据要求补全对应算法代码。

task3.txt为本题的数据文件，其内容为我国部分城市的经纬度坐标。task3.py为本题的代码文件。要求完成以下两个任务：

（1）根据经纬度坐标将所有数据利用Kmeans算法聚为10类；

（2）将聚类结果进行可视化（每一个城市表示为2维坐标平面的一个点，每一类中的点用同一种颜色，不同类的点采用不同颜色）。

🖉说明：将所补充代码源文件与运行结果截图放入答题卡对应位置。

**第四部分：深度学习算法设计及应用**（15分）

**🞏任务**：图片回归模型训练。

给定数据为一批人脸图片及每一张人脸对应的年龄标签数据。现要求把所有已给的数据当作训练集训练一个基于卷积神经网络的人脸年龄预测模型。

🖉要求与说明：

（1）本体旨在考察参赛者使用深度学习库解决具体计算机视觉问题的完整流程，不以测试集准确率作为评分标准；

（2）输入网络的数据尺寸统一缩放到64\*64\*3；

（3）使用Adam优化器；

（4）使用均方误差损失函数（MSE）；

（5）代码能正常运行，输出每一次迭代的损失值情况；

（6）使用深度学习库PyTorch或Tensorflow均可；

（7）卷积神经网络的结构如下：

|  |  |
| --- | --- |
| **层名称** | **参数配置** |
| 输入层 | 64\*64\*3 |
| 卷积层1 | 共8个3\*3的卷积核，步长为1，无padding |
| 非线性映射层1 | 非线性映射函数为Relu |
| 池化层1 | 最大值池化，核2\*2，步长为2 |
| 卷积层2 | 共16个3\*3的卷积核，步长为1，无padding |
| 非线性映射层2 | 非线性映射函数为Relu |
| 池化层2 | 最大值池化，核2\*2，步长为2 |
| 卷积层3 | 共32个3\*3的卷积核，步长为1，无padding |
| 非线性映射层3 | 非线性映射函数为Relu |
| 池化层3 | 最大值池化，核2\*2，步长为2 |
| 全连接层1 | 128个神经元+ Relu非线性映射 |
| Dropout | 激活概率为0.5 |
| 全连接层2 | 1个神经元+ Relu非线性映射 |

（8）需提交所有代码及成功训练运行截图。

**第五部分：人工智能技术综合应用**（20分）

**🞏任务**：根据房源信息，预测房屋价格。（数据为train.CSV, val.CSV, test.CSV）

房源信息包括：电梯情况|楼层|户型|区域|装修情况|面积|建筑时间|。

注：**部分信息有缺失**。训练集：验证集：测试集=17000：3000：3000

🖉要求与说明：

（1）根据训练集数据训练一个房屋出租价格预测模型。

（2）按解决问题的顺序**详细介绍每一步的目的**，**并给出相应代码，放入答题卡对应位置**。

（3）计算在验证集表现最好的模型在测试集数据的预测结果，将输出结果放入答题卡对应位置（**输出结果保存为txt格式并命名为task5result.txt，不能截图**）。

输出txt格式的数据每一行是测试集对应房源的预测价格（一定要符合格式要求，否则自动评分程序无法正确执行）

例：

800

120

615

…

（共3000行）

注：最终需要在规定位置提交的文件包括答题卡和task5result.txt。