《人工智能》

课程实验指导书

**目录**

实验一 搜索算法求解问题

实验二 约束满足问题

实验三 线性回归

实验四 逻辑回归

实验五 支持向量机

实验六 决策树

实验七 神经网络

附件：实验报告规范

**实验一 搜索算法求解问题**

本实验通过实验加深搜索算法求解问题的理解。建议使用python编程实现，并在Mindspore框架下实现。

一、实验目的

1. 掌握有信息搜索策略的算法思想；

2. 能够编程实现搜索算法；

3. 应用A\*搜索算法求解罗马尼亚问题。

二、实验内容及步骤

1. 创建搜索树；

2. 实现A\*搜索算法；

3. 使用编写的搜索算法代码求解罗马尼亚问题；

4. 分析算法的时间复杂度。

三、思考题

1. 宽度优先搜索，深度优先搜索，一致代价搜索，迭代加深的深度优先搜索算法哪种方法最优？

2. 贪婪最佳优先搜索和A\*搜索那种方法最优？

3. 分析比较无信息搜索策略和有信息搜索策略。

四、实验报告要求

1．说明实验的方法和步骤；

2. 对算法的原理进行说明；

3．给出算法源程序；

4．对实验结果进行分析。

**实验二 约束满足问题**

本实验通过实验加深约束满足问题的理解。建议使用python编程实现，并在Mindspore框架下实现。

一、实验目的

1. 求解约束满足问题；

2. 使用回溯搜索算法求解八皇后问题。

二、实验内容及步骤

实验内容：约束满足问题

三、实验报告要求

1．说明实验的方法和步骤；

2. 对算法的原理进行说明；

3．给出算法源程序；

4．对实验结果进行分析。

**实验三 线性回归**

本实验要求应用线性回归模型解决实际问题，通过实验加深对线性回归原理的理解。建议使用python编程实现，并在Mindspore框架下实现。

实验包括两部分：一元线性回归和多元线性回归。

**一、 一元线性回归**

题目：

应用一元线性回归预测移动餐车的利润。假设你是一家餐饮连锁店的CEO， 考虑在不同的城市开辟新店。该餐饮店已在许多城市拥有移动餐车，现有各个城市移动餐车的利润和城市人口的数据。这些数据将帮助你选择在哪个城市进行新店扩张。请按要求完成实验。

数据集：

文件ex1data1.txt为该实验部分的数据集，第一列表示城市人口（单位为万人），第二列表示该城市的移动餐车的利润（单位为万美元，若利润为负值，表示损失）。

步骤与要求：

1）在开始任务之前，进行数据的可视化对于了解数据特征是很有帮助的。请你导入数据并以人口为横坐标，利润为纵坐标画出散点图并观察数据分布特征。（建议：用python 的matplotlib）

2）将线性回归参数初始化为0，然后计算代价函数（cost function）并求出初始值。

3）使用线性回归的方法对数据集进行拟合，并用梯度下降法求解线性回归参数。（eg：迭代次数=1500，alpha=0.01）

4）画出数据的拟合图形。

5）预测人口数量为35000和70000时，利润为多少。

**二、 多元线性回归**

题目：

应用多元线性回归预测房价。假设你打算出售你的房子，你想知道房子的市

场价应该设为多少比较合适。一种方法就是收集最近的房屋销售信息并设计一个

房屋价格模型。请按要求完成实验。

数据集：

文件ex1data2.txt为该实验部分的数据集，第一列表示房屋的面积（平方英尺），第二列表示房间数目，第三列表示房屋价格。

步骤与要求：

1）导入数据，通过观察，容易发现房屋面积的大小约是房间数量的1000倍。当特征数量级不同时，对进行特征缩放能够使梯度下降更快地收敛。请对这两个特征进行归一化处理。

2）使用梯度下降法求解线性回归参数。尝试使用不同的alpha（学习率）进行实验，找到一个合适的alpha使算法快速收敛。思考alpha的大小对于算法性能的影响。

3）使用你认为最佳的alpha运行梯度下降法求出线性回归参数，然后预测房屋面积为1650平方英尺，房间数量为3时，房屋的价格。

**实验四 逻辑回归**

题目：

应用逻辑回归模型预测某学生能否被大学录取。假设你是某大学的系主任，你想根据两次考试的结果决定每个申请者的录取机会。现有以往申请者的历史数据，可以此作为训练集建立逻辑回归模型进行预测。（请按要求完成实验。建议使用python编程实现，并在Mindspore框架下实现。）

数据集：

文件ex2data1.txt为该实验的数据集，第一列、第二列分别表示申请者两次考试的成绩，第三列表示录取结果（1表示录取，0表示不录取）。

步骤与要求：

1）请导入数据并进行数据可视化，观察数据分布特征。（建议用python的matplotlib）

2）将逻辑回归参数初始化为0，然后计算代价函数（cost function）并求出初始值。

3）选择一种优化方法求解逻辑回归参数。

4）某学生两次考试成绩分别为45、85，预测其被录取的概率。

5）画出分类边界。

**实验五 支持向量机**

题目：

鸢尾花数据集（Iris data set）是模式识别中著名的数据集。本实验通过花萼（sepal）和花瓣（petal）的长和宽，建立SVM分类器来判断样本属于山鸢尾（Iris Setosa）、变色鸢尾（Iris Versicolor）还是维吉尼亚鸢尾（Iris Virginica）。请按要求完成实验。

数据集：

文件iris.txt为该实验的数据集，包含150个样本，对应数据集的每行数据。每行数据包含每个样本的四个特征（按顺序分别为花萼长度、花萼宽度、花瓣长度、花瓣宽度）和样本的类别信息（Iris Setosa、Iris Versicolor、Iris Virginica中的一种）。

实验要求：

（1）建立SVM分类器并用交叉验证法进行分析。

（2）利用PCA降维，将数据转化为二维，然后绘制出分类决策边界。

**实验六 决策树**

题目：

某连锁餐饮企业想了解周末和非周末对销量是否有很大影响，以及天气好坏、是否有促销活动对销量的影响，从而公司的辅助决策。现有单个门店的历史数据。请按要求完成实验。建议使用python 编程实现，并在Mindspore框架下实现。

数据集：

文件ex3data.xls 为该实验的数据集，第1-5 列分别表示序号、天气好坏、是否周末、是否有促销和销量高低。

实验要求：

选择ID3、C4.5、CART 三种常见决策树算法中的一种建立决策树模型，并画出决策树。（采用多种算法实现并进行算法的比较分析者，将获得更高分数）

**实验七 神经网络**

本实验包括两部分：交通运量预测和语音性别识别。

**一、交通运量预测**

题目：

公路运量主要包括公路客运量和公路货运量两个方面。据研究，某地区的公路运量主要与该地区的人口数量、机动车数量和公路面积有关。表1中给出了某地区20年的公路运量相关数据。

根据相关部门数据，该地区2010年和2011年的人口数量分别为73.39和75.55万人，机动车数量分别为3.9635和4.0975万辆，公路面积将分别为0.9880和1.0268万平方千米。

要求：

（1）请利用BP神经网络预测该地区2010年和2011年的公路客运量和公路货运量。

（2）请利用其他方法预测该地区2010年和2011年的公路客运量和公路货运量，并比较这两种方法的优缺点。

**表1 某地区20年公路运量数据**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 人口数量（万人） | 机动车数量（万辆） | 公路面积  （万平方米） | 公路客运量（万人） | 公路货运量（万吨） |
| 1990 | 20.55 | 0.6 | 0.09 | 5126 | 1237 |
| 1991 | 22.44 | 0.75 | 0.11 | 6217 | 1379 |
| 1992 | 25.37 | 0.85 | 0.11 | 7730 | 1385 |
| 1993 | 27.13 | 0.90 | 0.14 | 9145 | 1399 |
| 1994 | 29.45 | 1.05 | 0.20 | 10460 | 1663 |
| 1995 | 30.10 | 1.35 | 0.23 | 11387 | 1714 |
| 1996 | 30.96 | 1.45 | 0.23 | 12353 | 1834 |
| 1997 | 34.06 | 1.60 | 0.32 | 15750 | 4322 |
| 1998 | 36.42 | 1.70 | 0.32 | 18304 | 8132 |
| 1999 | 38.09 | 1.85 | 0.34 | 19836 | 8936 |
| 2000 | 39.13 | 2.15 | 0.36 | 21024 | 11099 |
| 2001 | 39.99 | 2.20 | 0.36 | 19490 | 11203 |

**二、语音性别识别（选做）**

题目：

根据声音样本来确定一个人的性别看似是一件很容易的事情。通常，人耳就可以很容易辨别出男性与女性声音的差异。那么，如何让计算机也能自动识别出男性或者女性的声音呢。

为了通过声音辨别性别，需要一个训练数据库。这个数据库由成千上万的男性和女性的声音样本建立起来，每个声音都以男性或女性的性别标记。

每个语音样本都存储为.WAV文件，然后对其进行预处理，测量声学信号的声学参数。预处理WAV文件的输出保存为一个CSV文件。

本实验，你将利用声音样本数据集，建立神经网络模型对声音进行分类。请按要求完成实验。

数据集：

文件train.csv为该实验的训练集，包含2536个声音样本，对应数据集的每行数据。每行数据包含每个样本的20个特征和样本的类别信息（男性或女性）。

文件test.csv为测试集，用于评估算法性能。格式与训练集相同。

实验要求：

（1）通过对训练集进行学习，建立神经网络模型并利用测试集测试算法分类的正确率(accuracy)。具体实验步骤应包括数据的可视化和预处理，模型建立与训练以及测试结果等。

（2）尝试使用其他机器学习方法进行实验，并作比较分析。

**附件：实验报告规范**

实验报告的开头应给出题目、班级、姓名、学号和完成日期，并至少包括以下几个内容：

1. 实验描述

简要描述实验中所用算法的基本思想（包括调参参数的选择等），以及如何处理数据集。

1. 实验及结果分析

（1） 开发语言及运行环境；

（2） 实验的具体步骤；

（3） 根据实验数据集，按实验要求给出相应的结果（截图）；

（4） 对实验结果进行简要分析。

1. 实验心得

对实验设计与实现过程进行总结和思考，记录实验过程的体会和心得。可包括：在实验中遇到的问题，又是如何解决的；体会调试的过程，总结编程中的收获。

1. 附录（可选）

带注释的源程序。

5、程序文件名的清单（必须写）

请列出所提交源程序电子版的程序文件名的清单，可根据实际情况补充对程序的说明，以便于实验验收。