# 第1章 人工智能概述

1.1 人工智能简介

1. 简答题

（1）人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

（2）略

（3）符号主义学派认为人类和机器的认知过程在本质上都是一种符号处理过程，人类思维过程总可以用某种符号来进行描述，其研究是以静态、顺序、串行的数字模型来处理智能，寻求知识的符号表征和计算。其有从定理机器证明、归结方法到非单调推理理论等一系列成就。

连接主义学派则是模拟发生在人类神经系统中的认知过程，提供一种完全不同于符号处理模型的认知神经研究范式，主张认知是相互连接的神经元相互作用的结果。其有归纳学习、深度学习、神经网络等代表性成果，直到今天仍在广泛使用。

行为主义学派认为智能是系统与环境的交互行为，是对外界复杂环境的一种适应。其有反馈控制模式及广义遗传算法等解题方法。

2. 思考题

略

1.2 人工智能研究内容

1. 简答题

（1）知识工程（Knowledge Engineering）是在计算机上建立专家系统的技术，图灵奖获得者、知识工程的建立者费根鲍姆（Feigenbaum）给出的知识工程定义是：将知识集成到计算机系统从而完成只有特定领域专家才能完成的复杂任务。其包括知识表示（knowledge representation）、知识获取（knowledge acquisition）、知识推理（knowledge reasoning）、知识集成（knowledge integration）和知识存储（knowledge storage）等多个活动。

（2）机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科，传统机器学习的研究方向主要包括决策树、随机森林、人工神经网络、贝叶斯学习等，而大数据时代的机器学习研究主要集中在数据转向以及数据信息处理能力等方面。

2. 思考题

略

1.3 人工智能技术与应用

1. 简答题

（1）图像（人脸）识别、语音识别、机器翻译、自动驾驶、下棋、气象预报、拣货机器人、交通调度等。

（2）略

2. 实践题

可以选取百度、腾讯、华为、阿里等云端人工智能系统进行介绍。具体答案略。

1.4 智能计算系统

1. 简答题

（1）智能计算系统包括了硬件和软件两大部分，硬件部分集成了通用CPU（Central Processing Unit，中央处理单元）和智能芯片的异构系统，软件部分包括面向开发者的智能计算编程环境，如编程框架和编程语言等。

（2）第一代智能计算系统，以面向符号主义的计算系统为代表，出现于1980年前后。第二代智能计算系统，以面向连接主义的计算系统为代表，自2010年前后发展至今。第二代智能计算系统形成于人工智能发展的第三次热潮，主要用于深度学习。未来出现的下一代智能计算系统或将成为强人工智能（通用人工智能）的物质载体。强人工智能是一种具有通用智能的机器的概念，该机器模仿人类的智能或行为，并具有学习和应用其智能来解决任何问题的能力。

（3）可以选配英伟达的Quadro RTX 2080、RTX 3080等高端GPU。

（4）此题没有标准答案，回答合理即可。例如，有的观点认为，未来随着技术的进一步发展，强人工智能的一些技术难点会被解决，强人工智能的出现只是时间问题；而也有一些专家呼吁不要开展强人工智能的研究。

2. 实践题

略

1.5 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1、D | 2、D | 3、A | 4、C | 5、B |
| 6、D | 7、D | 8、B | 9、C | 10、C |
| 11、C | 12、C | 13、A | 14、D | 15、B |
| 16、C | 17、D | 18、A | 19、D | 20、B |
| 21、D | 22、C | 23、A | 24、A | 25、C |

2. 填空题

1、图灵，1950，图灵测试

2、人工智能

3、连接主义

4、深度学习

5、符号主义，连接主义，行为主义

6、连接主义

7、模拟或实现人类的学习行为

8、神经网络及神经网络间的连接机制与学习算法

9、机器学习

10、智慧，智能应用

11、人工神经网络 或 神经网络

12、机器行为

13、盲目搜索，启发式搜索

14、深度学习

15、数据采集，数据清洗，特征工程，模型调优，模型验证

16、自动驾驶

17、组合优化

18、公有

# 第2章 人工智能体验

2.1 体验人工智能应用

1. 简答题

（1）如人脸识别（刷脸进站、刷脸支付、人脸解锁）、语音识别（语音助手、语音转文本）、文字识别（OCR）等。

2. 实践题

略

2.2 体验人工智能开发

1. 简答题

（1）创建“自动学习”项目；给素材标注好标签；设置训练参数，发布训练；项目发布上线。

（2）图像分类在很多领域有广泛应用，包括安防领域的人脸识别和智能视频分析等，交通领域的交通场景识别，互联网领域基于内容的图像检索和相册自动归类，医学领域的图像识别等。

（3）垃圾邮件过滤：提高基于统计的过滤方法的有效性，以达到过滤垃圾邮件的目的。

新闻分类：新闻文本的归档

舆情分析：检测网络中的是否违反政策法规，对大批量文本评论数据进行模型分类，情感分析。

2. 实践题

略

2.3 体验人工智能编程语言

1. 简答题

（1）数字、字符串、列表、元组、集合和字典等。

（2）Python支持三大类控制结构：顺序结构、分支结构以及循环结构。任何一个算法都可以使用这三种控制结构来设计完成。

2. 实践题

略

2.4 Anaconda开发环境

1. 简答题

（1）Anaconda Navigator、Anaconda Prompt、Jupyter Notebook、Spyder

（2）pip是Python默认的包管理器，一般来说新发布的工具包优先支持使用pip安装。而conda是一个与语言无关的跨平台环境管理器，可以安装任何语言的包，在安装时会检查当前环境下所有包之间的依赖关系，也可以进行虚拟环境的创建、删除等操作。

pip中的常用命令

安装包：pip install 包名

更新包：pip install 包名 --upgrade

卸载包：pip uninstall 包名

查询已经安装的包：pip list

查询已安装包的详细信息：pip show 包名

查询可以升级的包：pip list -o

查看pip的版本：pip -V

更新pip版本：pip install --upgrade pip

conda中的常用命令

安装包：conda install 包名

更新包：conda update 包名

卸载包：conda uninstall 包名

查询已安装的包：conda list

查看conda的版本：conda -V

更新conda版本：conda update conda

更新Anaconda版本：conda update anaconda

更新Pyhton版本：conda update python

（3）单击Python Shell窗口中的“File”，选择“New File”创建Python文件；在文件中写入代码；单击该Python文件窗口中的“File”，选择“Save”，在弹出的窗口中选择保存路径，输入文件名，单击“保存”，保存当前代码在指定路径下的<文件名>.py文件中；单击<文件名>.py文件窗口中的“Run”，选择“Run Module”可执行该文件中的代码，在Python Shell窗口中可以看到执行结果。

（4）在Jupyter Notebook的主界面单击右上角的下拉列表“New”，选择“Python3”，可以创建一个新的“ipynb”格式笔记，浏览器在新标签页中显示该笔记本。在笔记本的第一行输入代码，单击工具栏中的“运行”按钮执行该代码块，代码块下方会显示代码的执行结果。Jupyter Notebook也可以作为文档书写工具使用。选择一个输入框，单击工具栏中的“代码”下拉框，选择“Markdown”（或“标记”），此时在输入框的格式由“代码”转为“Markdown”（或“标记”），用户就可以在输入框中写入说明文字。然后，单击工具栏中的“运行”按钮后，该处说明文字将以文档的形式呈现。

2. 实践题

略

2.5 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1、B | 2、D | 3、C | 4、A | 5、D |
| 6、A | 7、B | 8、A | 9、A | 10、D |

2. 填空题

1、图像

2、自然语言处理

3、对象存储服务 或 OBS

4、积木

5、分支

6、IDLE

3. 综合实践

略

# 第3章 人工智能编程语言

3.1 Python语言及基本语法

1. 简答题

（1）ABFHI

（2）ADF

（3）AC

（4）"#"开始的单行注释；多行字符串常量（三个单引号开始三个单引号结束，或者三个双引号开始三个双引号结束）表示的多行注释。

2. 实践题

（1）

>>> type(128)

<class 'int'>

>>> type(128.0)

<class 'float'>

>>> type(2+5j)

<class 'complex'>

>>> type(True)

<class 'bool'>

>>> type((1,2,3,4,5))

<class 'tuple'>

>>> type({1,2,2,3,4})

<class 'set'>

>>> type({"a":94,"c":96,"A":65,"C":67})

<class 'dict'>

（2）

>>> height=float(input("请输入你的身高(cm)："))

请输入你的身高(cm)：156

>>> weight=float(input("请输入你的体重(kg)："))

请输入你的体重(kg)：46.5

>>> BMI=weight/pow(height/100,2)

>>> print("你的BMI值为：{:.1f}".format(BMI))

你的BMI值为：19.1

3.2 组合数据对象

1. 简答题

（1）'0'<=ch<='9'

（2）

方法一：

>>> heading="this is a sample!"

>>> heading=heading.split()

>>> for i in range(len(heading)):

         heading[i]=heading[i].capitalize()

>>> heading=' '.join(heading)

>>> heading

'This Is A Sample!'

方法二：

>>> heading="this is a sample!"

>>> heading.title()

'This Is A Sample!'

（3）S2=S1[::-1]

（4）设置字符串的sort()方法、或内置函数sorted()函数的参数key=len

（5）将每一个项目的运动员名单列表连接（+）为一个总列表，转化为集合去重复后再转化回列表。

L=list(set(L1+L2+L3+…))

2. 实践题

（1）HelloPython；True；HelloPythonPythonPython

（2）

>>> str="Hello,Python World!"

>>> str[0]

'H'

>>> str[5]

','

>>> str[6:-7]

'Python'

>>> str[::-1]

'!dlroW nohtyP,olleH'

（3）'PYTHON STRING'；'python string'；'Python String'；10；'Python Strgni'；['Python', 'String']

（4）s1[:7]；s1[:3]+s2[:3]；s1[5:7]\*3；(s1+' '+s2).replace('a', '@')

（5）' 001001'；'Li Si'；tuple(元组)；4

（6）' 001001'；'Li Si'；列表(list)

（7）把两个列表中的元素连接成一个列表；把t2列表中的第一个元素替换成空列表(删除列表中第一个元素)

（8）{4:10, '1':20}

（9）r={p} 和 w={s}

（10）0

3.3 程序控制结构

1. 简答题

（1）顺序结构、选择结构、循环结构

（2）for循环为遍历循环，一般用在循环次数已知或者能进行迭代遍历的数据访问中，while循环为条件循环，一般用在由条件控制的循环结构中。

（3）break跳出并结束当前循环，执行循环后的语句，continue结束当次循环，继续执行后续次数循环。break和continue可与for 和while搭配使用。

2. 实践题

见实验答案

3.4 模块和函数

1. 简答题

（1）

方法一：导入一个模块

导入标准库：import  <模块名>

使用函数：<模块名>.<函数>(<参数>)

举例：求8的平方根

>>> import math

>>> math.sqrt(8)

2.8284271247461903

方法二：导入一个模块中所有对象

导入标准库：from <库名> import \*

使用函数： <函数>(<参数>)

举例：求8的平方根

>>> from math import \*

>>> sqrt(8)

2.8284271247461903

方法三：导入模块中的指定对象

举例：求8的平方根

>>> from math import sqrt

>>> sqrt(8)

2.8284271247461903

（2）

def和lambda都是Python中定义函数的方式。lambda是没有函数名字的匿名函数定义，返回函数对象，常用来定义临时使用的一个表达式(单行)就能表示结果的函数。一些复杂的函数只能用def定义，def定义函数更具有通用性。判断一个数是否是偶数的函数定义：

def定义函数：

def isEven(x):

    if x%2==0:

        return True

    else:

        return False

lambda定义函数：

isEven=lambda  x:x%2==0

有函数名的函数调用时，通过函数名及实参完成调用，而lambda函数调用时通过返回的函数对象实现的调用。

2. 实践题

见实验答案

3.5 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1、C | 2、A | 3、D | 4、B | 5、D |
| 6、D | 7、B | 8、B | 9、A | 10、C |
| 11、D | 12、D | 13、C | 14、A | 15、D |
| 16、B | 17、C | 18、B | 19、B | 20、B |
| 21、C | 22、C | 23、B | 24、C | 25、B |
| 26、A | 27、C | 28、D | 29、D | 30、C |

2. 综合实践

见实验答案

# 第4章 人工智能数据处理

4.1 NumPy数据类型

1. 简答题

（1）NumPy数组要求成员的数据类型一致，从而提高运算速度；Python列表对象的成员数据类型可以不同。

（2）使用索引返回具体位置的元素，使用切片返回指定范围的元素。

（3）linspace函数用于创建指定区间内等间隔一维数组，uniform函数用于创建指定区间内随机服从均匀分布的一维数组。

2. 实践题

见实验答案

4.2 Pandas数据类型

1. 简答题

（1）Series适用于一维数据，DataFrame适用于表格数据。

（2）参数的类型不同，函数名不同。

（3）由于DataFrame适用于表格数据，因此DataFrame更加常用。

2. 实践题

见实验答案

4.3 表格数据处理

1. 简答题

（1）value的类型可以是常量、字典、Series或DataFrame。

（2）保留重复行中第一次出现的行和保留重复行中最后一次出现的行。

（3）两个函数都是用于删除缺失值所在的行或列的，同时都默认不会改变原始对象；不同之处在于函数参数不同。

2. 实践题

见实验答案

4.4 数据可视化

1. 简答题

（1）散点图通常用于研究两组变量之间的数据相互关系；

   折线图能够显示数据在一个连续的时间间隔上的变化#从而反映事物随时间或有序类别而变化的趋势；

   条形图用来进行各类别数据大小的比较；

   箱型图用作显示一组数据分散情况资料的统计图；

   饼图用于描述量、频率或百分比之间的相对关系。

（2）依具体情况和个人喜好而定

（3）常见颜色有蓝，绿，红，黄，黑等。

2. 实践题

见实验答案

4.5 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1.B | 2.A | 3.D | 4.A | 5.C |
| 6.A | 7.D | 8.D | 9.C | 10.A |
| 11.B | 12.D | 13.A | 14.A | 15.C |
| 16.C |  |  |  |  |

2. 填空题

1、astype()

2、random.normal()

3、array.size

4、append()

5、缺失数据的清除，缺失数据的填充

6、‘first’，‘last’，False

7、行/列的全部值都为缺失值才会进行移除

8、jieba

9、grid()

10、scatter()

3. 综合实践

见实验答案

# 第5章 机器学习

5.1 人工智能与机器学习

1. 简答题

（1）简述模型训练中训练集、测试集、验证集的含义。

训练集是训练机器学习算法的数据集；测试集是用来评估经训练后的模型性能的数据集；验证集是用来微调模型超参数的数据集。

（2）简述预测值与真实值之间的四种关系及其含义。

四种关系如下：

真正类（TP，True Positive）：预测正确，预测该样本为正类，真实类别为正类；

假正类（FP，False Positive）：预测错误，预测该样本为正类，真实类别为负类；

假负类（FN，False Negative）：预测错误，预测该样本为负类，真实类别为正类；

真负类（TN，True Negative）：预测正确，预测该样本为负类，真实类别为负类。

（3）分别简述数据的归一化、标准化、正则化的含义。

归一化：同一数据集中，不同列的数据往往有着完全不同的含义，数值大小差异很大，可能会影响数据处理的最终结果，因此常常需要把每列数据都映射到0-1范围之内处理，即归一化。

标准化：对数据进行标准化处理的函数，包括Z-score标准化、稀疏数据标准化和带离群值的标准化。其中，Z-Score方法可以在大多数类型的数据上得到较好的应用，标准化后得到的数据是以0为均值，1为方差的正态分布。但由于它是一种中心化的方法，将会对原始数据的分布结构产生改变。

正则化：将单个样本缩放到单位范数。

（4）简述什么是机器学习的过拟合现象，过拟合产生的原因。

模型在训练集上表现很好，但在测试集中训练效果较差，导致的原因是模型过分的考虑训练集中的已知数据自身特性，导致对新数据的训练效果变差，表明数据集的变化对模型性能的影响。

2. 实践题

见实验答案

5.2 分类

1. 简答题

（1）① 分类算法是有监督学习，聚类是无监督学习。

② 分类算法的样本数据集已被标记，聚类算法的样本数据集无需标记

（2）常见的分类算法有K最近邻算法：KNN（K-Nearest Neighbor）、朴素贝叶斯分类算法：NBC（Naive Bayesian Classifier）、逻辑回归算法：LR（Logistic Regression）、决策树算法：Decision Tree、支持向量机算法：SVM（Support Vector Machine）、人工神经网络：ANN（Artificial Neural Network）

2. 实践题

见实验答案

5.3 回归

1. 简答题

（1）回归可以看作是研究一个或多个因变量 与另一个或多个自变量之间的依存关系，用自变量的值来估计或预测因变量的总体平均值。

（2）线性回归是利用称为线性回归方程的最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。当函数为参数未知的线性函数时，称为线性回归。线性回归使用最佳的拟合线（回归线）在因变量Y和自变量X间建立一种关系。线性回归模型是线性预测函数，模型参数通过样本数据来估计。在这种技术中，自变量可以是连续的也可以是离散的。

（3）多元线性回归用来分析多个输入变量共同影响输出变量的问题。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多元线性回归分析。

（4）线性回归分析过程一般包括：确定自变量和因变量、建立预测模型、变量间的相关性检验、模型的评估和检验、利用模型进行预测等阶段。

2. 实践题

见实验答案

5.4 聚类

1. 简答题

（1）聚类算法中常用的距离公式有欧式距离、曼哈顿距离。

（2）衡量聚类算法性能主要是通过簇间距离和簇内距离来衡量，通常有调整兰德系ARI、调整的互信息指数AMI等。

（3）K-Means算法结束算法迭代的条件通常有质心是否发生更新、迭代次数。

2. 实践题

见实验答案

5.5 降维

1. 简答题

（1）降维的主要动机是去除冗余信息和噪声，压缩数据，加速计算，便于表示和可视化。降维的负面影响是会造成信息的损失、转化后的特征难以解释。

（2）PCA可以用来给高维度非线形数据集降维。

（3）可以使用精确率(Precision)、召回率(Recall)、F1分数(F1 Score)、准确率(Accuracy)和训练耗时来评估降维算法的性能。

（4）正交投影矩阵可以保证投影到的低维特征空间的每个维度是不相关的，降低特征之间的耦合性；非正交投影矩阵保留了属性间的联系。

2. 实践题

见实验答案

5.6 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1.A | 2.D | 3.C | 4.A | 5.B |
| 6.C | 7.B | 8.C | 9.B | 10.B |
| 11.D | 12.A | 13.A | 14.B | 15.D |
| 16.B | 17.D | 18.D | 19.B | 20.D |
| 21.C | 22.A | 23.A | 24.A | 25.A |
| 26.A | 27.D | 28.D | 29.C | 30.B |
| 31.A |  |  |  |  |

2. 填空题

1、监督学习，半监督学习，强化学习

2、MAE 或 平均绝对误差，MSE 或 均方误差，RMSE 或 均方根误差

3、训练集测试集验证集

4、5

5、相似性

6、距离

7、有监督

8、准确率，精确率，召回率

9、高

10、甲

11、多元

12、截距

13、回归

14、分类

15、监督

16、正交特征主成分

17、PCA、LDA、MDS、LLE等

18、冗余信息 或 冗余数据，计算

3. 综合实践

见实验答案

# 第6章 深度学习

6.1 TensorFlow游乐场

1. 简答题

（1）TensorFlow游乐场主要包括DATA（数据）、FEATURES（特征）、HIDDEN LAYERS（隐藏层）、OUTPUT（输出）和参数设置五个区域。

（2）层与层之间的连线表明权重值的大小，橙色代表权值为负值，蓝色代表正值。

（3）增加输入特征的数量，增加隐藏层的层数和每层神经元的数目。

2. 实践题

（1）略

（2）略

（3）略

6.2 神经网络基本原理

1. 简答题

（1）神经网络中常用的激活函数有Sigmoid函数、tanh函数和ReLU函数等

（2）Sigmoid函数并非以原点为中心，当远离坐标原点时，f(x)的梯度会非常小，接近于零，从而出现梯度弥散现象，不利于神经网络模型中权重的更新。

tanh函数是双曲正切函数，与Sigmoid函数的图形类似，不同之处是tanh函数的值域为(-1,1)，函数是以原点为中心。tanh函数作为隐藏层的激活函数要比Sigmoid函数效果好一些，而对于二分类问题输出层一般选择Sigmoid函数，因为Sigmoid函数的取值范围为(0,1)，可以理解为计算分类的概率。

（3）当x>0时，ReLU函数的梯度恒为1，不仅增强了神经网络的梯度下降法的运算速度，还克服了tanh函数和Sigmoid函数在远离原点位置的梯度弥散问题，这也是ReLU函数的一大特点。

（4）BP算法的反向传播是根据正向传播后得到的误差值（预测值与真实值之差），因此为监督学习的算法。

2. 实践题

（1）y1=f(4\*(-0.1)+4\*0.6)=f(2)

y2=f(5\*0.4+5\*(-0.2))=f(1)

6.3 神经网络构建

1. 简答题

（1）Keras是一个对用户友好、用Python编写的高度模块化的神经网络库，并可以采用TensorFlow、CNTK或 Theano 作为后端对深度学习进行底层运算。TensorFlow 2.0将Keras作为其高阶API已经集成其中，通过tf.keras调用。

（2）利用Keras构建神经网络模型时，主要涉及的步骤为：载入数据、数据预处理、构建Sequential模型、利用compile函数进行编译、利用fit函数训练模型、模型的评估和对新数据的预测。

（3）增加隐藏层神经元的数目；增加隐藏层的层数。

2. 实践题

（1）参考6.3.1小节内容。

（2）见实验答案

（3）见实验答案

6.4 卷积神经网络

1. 简答题

（1）传统的神经网络模型是由输入层、隐藏层和输出层组成，且每层都是由多个神经元构成（模型输入也是经reshape()函数转换的一维向量）。卷积神经网络模型与其有显著不同： 该模型主要是由卷积层、池化层、全连接层等构成；

（2）卷积层，顾名思义，是对输入的图像进行“卷积”运算，获取图像的特征。浅层的卷积层主要学习图像的低级基础特征，随着层数的增加，深层的卷积层学习到更高级的图像特征。

池化层在卷积神经网络中主要是对图像进行下采样降低图像的尺寸，即卷积层输出的特征向量维数，从而降低过拟合现象和减少需要计算的参数。

全连接层对应6.3节简单版程序的神经网络结构，包括输入层、隐藏层和输出层。进入全连接层时需要将卷积层的多维数据利用Flatten()函数进行扁平化变为一维向量，最后输出层可以是单神经元(二分类)或者多个神经元(多分类)。

（3）利用Keras构建卷积神经网络模型时主要涉及的步骤为，载入数据，对该数据预处理，构建Sequential模型，构建神经网络和全连接层，利用compile函数进行编译，利用fit函数训练模型，最后进行模型的评估和对新数据的预测。

2. 实践题

见实验答案

6.5 综合练习

1. 选择题

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1、A | 2、D | 3、D | 4、A | 5、B |
| 6、D | 7、B | 8、A | 9、D | 10、A |
| 11、D | 12、A | 13、A | 14、C | 15、A |
| 16、C | 17、D | 18、A | 19、B |  |

2. 填空题

1、节点，神经元权重值

2、回归

3、训练次数，学习率，激活函数

4、正向传播，反向传播

5、Model 或 模型

6、API

7、CNN，RNN

8、输入层，隐含层，输出层

9、卷积层，池化层，全连接层

10、compilefit

3. 综合实践

（1）略

（2）y=0.8\*1.1-0.1\*0.6+2\*0.3=1.42

（3）见实验答案

（4）见实验答案

（5）见实验答案

（6）见实验答案