**人工智能算法复习题**

**机器学习**

**基础概念**

1、在Regression模型中，加入L1或L2正则化，会产生什么效果？

可以做特征选择，防止过拟合

2、向量x=[1, 2, -3, 4, -5，0]的L1范数是多少？

绝对值相加

3、对于模型的高偏差，有什么办法可以降低？（B）

A、减少特征 B、增加特征 C、增加数据点 D、B和C

4、理解分类算法的准确率、召回率和F1值

准确率：是指检索出相关文档数与检索出的文档总数的比率，衡量的是检索系统归档查准率  
𝑇𝑃/(𝑇𝑃+𝐹𝑃)

召回率：是指检索出的相关文档数与文档库中所有的相关文档数的比率，衡量的是检索系统归档查全率  
𝑇𝑃/𝑃

F1值：为了解决准确率和召回率冲突的问题，引入了F1值  
(2×𝑝𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛×𝑟𝑒𝑐𝑎𝑙𝑙)/(𝑝𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛+𝑟𝑒𝑐𝑎𝑙𝑙)

5、以下哪些是梯度下降算法？（ABC）

A、BGD B、MBGD C、SGD D、AGD

6、梯度下降算法的正确步骤是？（43152）

1.计算预测值和真实值之间的误差

2.迭代更新，直到找到最佳权重

3.把输入传入网络，得到输出值

4.初始化随机权重和偏差

5.对每一个产生误差的神经元，改变相应的权重值以减小误差

8、Pandas包含的两种主要数据类型是？

9、以下关于机器学习描述正确的是哪一项？（A ）

A、深度学习是机器学习的一个分支

B、深度学习和机器学习是互相包含的的关系

C、深度学习与机器学习同属于人工智能但相互之间没有关系

D、以上都不对

10、当函数在变量空间的某一点处，沿着哪一个方向有最大的变化率？（C）

A、导数 B、方向导数 C、梯度 D、最小方向导数

11、以下表述正确的是？（C）

A、机器学习模型在训练集上的精准度越高，则模型的性能越好

B、增加模型的复杂度，总能减小测试样本误差、

C、增加模型的复杂度，总能减小训练样本误差

D、以上说法都不对

12、假设需要改变参数来最小化代价函数，可以使用以下哪项技术？（D）

A、网格搜索 B、随机搜索 C、Bayesian优化 D、以上任意一种

13、以下选项中对泊松分布于二项分布的关系描述正确的是？

泊松分布是二项分布当n很大p很小时的近似计算

14、以下关于监督式学习的说法正确的是？（C）

A、监督式不需要标签就可以训练

B、监督式学习是一种基于规则的算法

C、决策树是一种监督式学习

D、监督式学习不可以使用交叉验证进行训练

15、以下正则表达式中，哪个选项用于匹配除换行符外的任意字符？（A）

A、. B、^ C、$ D、？

16、首次提出“人工智能”是在哪一年？

1956

17、特征选择的方法有哪些？（ABCD）

A、方差选择法 B、相关系数选择法 C、卡方检验法 D、互信息法

**判断题**

1、在回归分析中，岭回归和Lasso模型计算量小，且Lasso回归可以实现特征选择。（正确）

2、L2正则化比L1正则化产生更加稀疏的模型。（错）

3、精确率计算公式  
𝑇𝑃/(𝑇𝑃+𝐹𝑃)

4、小批量梯度下降中的小批量是指在每次参数迭代时都是用从特征中抽出一部分进行训练。（错）

5、欠拟合的原因是模型复杂度过高，训练数据过少，训练误差小，测试误差大。（错）

6、召回率的计算公式  
𝑇𝑃/𝑃

7、L2正则化能防止过拟合，提升模型的泛化能力，但L1做不到这点。（错）

8、回归问题和分类问题都有可能发生过拟合。（正确）

9、当训练数据较少时更容易发生过拟合。（错）

10、梯度下降有时会陷于局部极小值，但EM算法不会。（错）

11、在核回归中，最影响回归的过拟合和欠拟合之间平衡的参数为核函数的宽度。（正确）

12、一般来说，回归不用在分类问题上，但也有特殊情况，比如logistic回归可以用来解决0/1分类问题。（正确）

**决策树和集成学习**

1、以下关于决策树的说法正确的是哪一项？（B）

A、ID3决策树可以处理连续值的特征

B、C4.5决策树可以处理连续值的特征

C、CART决策树不能处理空值

D、C4.5决策树不能处理空值

2、以下哪些选项是决策树用于划分节点的依据？(AC)

A、Gini系数 B、ID3 C、信息熵 D、CART

3、Bagging主要通过什么来降低方差？（B）

A、多次迭代 B、平均 C、更多数据 D、优化算法

**判断题**

1、Boosting算法的一个优点是不会过拟合。（错）

2、在AdaBoost算法中，所有被错分的样本的权重会被增大。（正确）

3、随机森林算法能够用来做特征选择。（正确）

**支持向量机SVM**

1、svm通过什么方法扩展到非线性边界？（B）

A、多元函数 B、核函数 C、多个特征 D、多变量

2、二分问题时，当训练集是线性可分的情况下，svm可以用什么进行分类？（B）

A、软间隔 B、硬间隔 C、核函数 D、张量

3、以下关于SVM算法说法错误的是哪一项？（A）

A、SVM算法不能用于非线性数据集的分类

B、SVM算法可以使用多种核函数

C、支持向量是距离超平面最近的向量点

D、SVM可以被用于解决分类问题

**聚类和降维**

1、alllomerative clustering算法的主要分割思想是什么？（A）

A、自下而上 B、聚类 C、自上而下 D、回归

2、K-means Clustering算法的k值如何选取的？（B）

A、随机 B、人工 C、计算机自动识别 D、已知

3、常见的聚类算法有哪些？（ABCD）

A、K-means B、层次聚类 C、密度聚类 D、谱聚类

**关联规则**

1、某数据库中有三次交易，分别为物品A，B；物品B，D；物品A，D；请问物品A的支持度计数是多少？

2

2、某数据库中有三次交易，分别为物品A，B；物品B，D；物品A，D；请问物品A->B的置信度是多少？

0.5

3、Apriori算法通过什么指标筛选频繁项集得出有效的关联规则？（B）

A、C值 B、置信度 C、支持度 D、alpha

## 深度学习

1、对于图像识别问题，用什么神经网络模型结构更适合解决此类问题？

卷积神经网络

2、对于分类任务，如果我们不是将神经网络中的权重随机初始化，而是将所有权重重设为零。则会发生什么？

神经网络模型可以训练，但所有的神经元最终将识别同样的事情。

3、若要运用RNN构建情感分析模型，应采用哪种连接方式？（C）

A、one to one B、one to many C、many to one D、many to many

4、在一个神经网络里，知道每一个神经元的权重和偏差很重要。如果以某种方法知道了神经元准确的权重和偏差，就可以你和任何函数。实现这个最佳的办法是什么？（C）

A、随机赋值，祈祷它们是正确的

B、搜索所有权重和偏差的组合，直到得到最佳值

C、赋予一个初始值，通过检查跟最佳值的差值，然后迭代更新权重

D、以上都不对

5、在神经网络中，以下哪种技术用于解决过拟合？（D）

A、Dropout B、正则化 C、批规范化 D、所有

6、在生成对抗网络中，带有标签的数据应该被放入哪里？（B）

A、作为判别模型的输出值

B、作为判别模型的输入值

C、作为生成模型的输入值

D、作为生成模型的输出值

7、神经网络属于哪个学派？（B）

A、符号主义

B、连接主义

C、行为主义

D、都不是

8、人工智能中，感知智能的技术方向分为以下哪几种？（BCD）

A、知识图谱 B、计算机视觉 C、自然语言处理 D、智能语音处理

9、哪些属于深度学习算法的激活函数？（ABC）

A、sigmoid B、Relu函数 C、Tanh函数 D、Sin函数

10、以下哪些属于LSTM的网络结构？（ABC）

A、输入门 B、遗忘门 C、输出门 D、记忆门

11、关于循环神经网络的描述，正确的有？（ABC）

A、循环神经网络可以简写为RNN

B、循环神经网络可以根据时间轴展开

C、LSTM也是一种循环神经网络

D、LSTM无法解决梯度消失的问题

12、关于卷积神经网络池化层的描述，正确的是哪些？（BCD）

A、池化操作也是如同卷积一样采用扫描窗口操作

B、池化层可以起到降维的作用

C、常用的池化有最大池化和平均池化

D、经过池化的特征图像变小了

13、关于反向传播，以下说法正确的是哪些？（ABD）

A、反向传播指的是误差通过网络反向传播

B、反向传播会经过激活函数

C、反向传播只能在前馈神经网络中运用

D、反向传播可以结合梯度下降算法更新网络权重

14、关于梯度下降以下说法错误的是哪些？（ABD）

A、使用梯度下降时，学习率应该保持一致

B、全局梯度下降比随机梯度下降速度更快

C、随机梯度下降对噪声点敏感

D、Adam不属于梯度下降算法

15、以下关于神经网络的说法，正确的有哪些？（ACD）

A、单层感知器的局限在于不能解决异或问题

B、随着神经网络隐藏层数的增加，模型的分类能力逐步减弱

C、前馈神经网络同一层神经元之间不存在联系

D、前馈神经网络可用有向无环图表示

16、关于卷积神经网络的说法，正确的有哪些？（ACD）

A、卷积神经网络包含卷积层，池化层，全连接层

B、卷积核不可以用来提取图片全局特征

C、常见的池化层有最大池化和平均池化

D、在处理图片时，是以扫描窗口的方式对图像做卷积

17、以下哪些激活函数容易产生梯度消失问题？（AD）

A、Sigmoid B、Softplus C、Relu D、Tanh

18、深度学习中，哪些步骤是由模型自动完成的？ ( BCD )

A、分析定位任务

B、模型训练

C、特征提取

D、特征选择

19、深度学习中常用的损失函数有哪些？（CD）

A、L1损失函数

B、自下降损失函数

C、均方误差损失函数

D、交叉熵损失函数

20、深度学习模型训练中，常见的优化器有哪些？（ABCD）

A、SGD

B、Adagrad

C、Adam

D、动量优化器

### 判断题

1、对于计算机视觉任务来说，使用残差网络能够解决随着网络深度的增加而产生的梯度消失问题。（正确）

2、softmax激活函数的作用是减少计算量和防止梯度消失。（错）

3、修正线性单元（Relu）使得神经网络中引入了非线性。（错）

4、构建深度学习模型时，tf.keras.Sequential中封装了常用的网络层，包括CNN和RNN等。（错）

5、在卷积神经网络中添加池化层（pooling layer）时，能一定程度上解决图像不变性问题。（错）

6、标准RNN模型会出现长时间传输记忆力的信息衰减问题。（正确）