

机器学习模拟试卷一

一、填空题（每空 1 分，共 20 分）

1. 机器学习是一门利用_____来模拟和实现人类学习过程的学科。
2. 根据学习过程是否提供标签，机器学习可分为监督学习、无监督学习和_____学习。
3. 经验误差是指模型在_____上的误差，泛化误差是指模型在_____上的误差。
4. 解决过拟合的常见方法有增加训练数据、降维、_____。
5. 在模型评估中，_____法将数据集随机划分为训练集和测试集。
6. 分类任务中常用的性能度量有准确率、查准率、查全率和_____。
7. 逻辑回归使用的激活函数是_____，其输出值范围为_____。
8. 朴素贝叶斯法的基本假设是_____。
9. KNN 算法中， k 的含义是_____。
10. 决策树 ID3 算法使用_____作为划分标准，C4.5 使用_____。
11. 集成学习中，Bagging 主要用于降低_____，Boosting 主要用于降低_____。
12. SVM 中处理线性不可分问题的技术称为_____。
13. K-means 聚类需要预先指定的参数是_____。
14. PCA 是一种_____方法，其目标是找到数据中方差最大的方向。
15. 神经网络中，用于缓解梯度消失问题的激活函数是_____。
16. 梯度下降法中，每次使用一个样本更新参数的是_____梯度下降。

二、判断题（每题 1 分，共 20 分）

(正确打“√”，错误打“×”)

1. 机器学习中，数据量越大模型性能一定越好。 ()
2. 过拟合是指模型在训练集上表现好，在测试集上表现差。 ()
3. 逻辑回归只能用于二分类问题。 ()
4. 梯度下降法一定能找到全局最优解。 ()
5. 数据归一化会改变数据的分布。 ()
6. L1 正则化可以使部分权重变为 0。 ()
7. 朴素贝叶斯是生成模型。 ()
8. 交叉验证只适用于小数据集。 ()
9. KNN 算法对异常值不敏感。 ()
10. 决策树 ID3 算法可以处理连续特征。 ()
11. 随机森林不容易过拟合。 ()
12. SVM 只能处理线性可分数据。 ()
13. 聚类是无监督学习任务。 ()
14. PCA 降维后信息完全无损失。 ()
15. 关联规则 Apriori 算法效率高于 FP-Growth 算法。 ()
16. 感知机可以解决非线性分类问题。 ()
17. 神经网络中 ReLU 函数可能导致神经元“死亡”。 ()
18. 批梯度下降的更新频率高于随机梯度下降。 ()
19. 查准率和查全率是正相关关系。 ()
20. 集成学习可以同时降低方差和偏差。 ()

三、简答题（每题 6 分，共 30 分）

1. 什么是欠拟合和过拟合？各举一种解决方法。

2. 写出逻辑回归的代价函数，并解释其含义。

3. 简述 K-means 聚类算法的基本步骤。

4. 比较 Bagging 和 Boosting 的异同。

5. 什么是核技巧？SVM 中常用的核函数有哪些？

四、综合题（每题 10 分，共 30 分）

1. （推导题）请推导逻辑回归的梯度下降更新公式。已知逻辑回归的假设函数为：

$$h(x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

损失函数为：

$$L(\hat{y}, y) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

整体代价函数定义为：

$$J(w) = \frac{1}{m} \sum_{i=1}^m L(h(x^{(i)}), y^{(i)})$$

请推导参数 w_j 的梯度下降更新公式：

$$w_j := w_j - \alpha \frac{\partial J(w)}{\partial w_j}$$

其中， α 为学习率。

2. 给定一个二分类任务，训练集中正负样本数量极不均衡（正样本极少），请：
列举两种处理不平衡数据的方法，并说明原理；
如果使用逻辑回归模型，你会如何调整模型以应对该问题？

3. 假设你正在构建一个垃圾邮件分类系统，请：
选择一种合适的机器学习算法，并说明理由；
描述该算法的建模流程；
说明你将如何评估模型的性能。

机器学习模拟试卷二

一、填空题（每空 1 分，共 20 分）

1. 在监督学习中，如果标签是连续值，则该任务属于_____；如果标签是离散值，则属于_____。
2. 模型评估中，_____误差反映了模型在训练集上的表现，_____误差反映了模型在新样本上的表现。
3. 解决欠拟合的常见方法有增加特征、_____和_____。
4. 交叉验证法通过_____划分数据集，常用的有 k 折交叉验证。
5. 在二分类问题中，混淆矩阵的四个基本元素是 TP、FP、_____ 和 _____。
6. 逻辑回归中，sigmoid 函数的导数公式为

$$g'(z) = g(z)(1 - g(z))$$

其中

$$g(z) = \text{_____}$$

7. 朴素贝叶斯分类器中，为了避免概率为 0，通常使用_____平滑技术。
8. KNN 算法中，常用的距离度量包括欧氏距离、曼哈顿距离和_____距离。
9. 决策树中，CART 算法使用_____指数作为分类树的划分标准。
10. 随机森林通过_____采样构建多棵决策树，并通过投票或平均进行预测。
11. AdaBoost 算法通过调整_____权重来提升错分样本的重要性。
12. SVM 中，_____参数控制对误分类样本的惩罚程度。
13. K-means 聚类中，肘部法则用于选择最佳的_____值。
14. DBSCAN 聚类基于样本密度，将样本分为核心点、_____和噪声点。
15. PCA 降维的目标是最大化投影后的_____。
16. 关联规则 Apriori 算法基于_____原理，即频繁项集的子集也是频繁的。
17. 神经网络中，_____函数可以有效缓解梯度消失问题，并加速收敛。
18. 梯度下降法中，_____梯度下降在每次更新时使用一个批次的样本。

19. 在模型选择中，_____误差用于选择超参数，_____误差用于最终评估模型性能。

20. 正则化项中，L1 正则化倾向于产生_____模型。

二、判断题（每题 1 分，共 20 分）

(正确打“√”，错误打“×”)

1. 监督学习必须使用带标签的数据进行训练。 ()
2. 过拟合无法避免，只能通过正则化等手段缓解。 ()
3. 逻辑回归的损失函数是均方误差。 ()
4. 归一化和标准化都会改变数据的原始分布。 ()
5. L2 正则化可以使部分权重恰好为 0。 ()
6. 朴素贝叶斯假设所有特征之间相互独立。 ()
7. KNN 算法中， k 值越大，模型越复杂，越容易过拟合。 ()
8. 决策树 ID3 算法可以处理连续型特征。 ()
9. 随机森林中每棵树的训练样本都是相同的。 ()
10. AdaBoost 中，分类误差率小的弱分类器权重较大。 ()
11. SVM 只能用于二分类问题。 ()
12. 核函数可以将低维数据映射到高维空间，使其线性可分。 ()
13. 聚类分析中，K-means 对异常值不敏感。 ()
14. 层次聚类可以生成树状的聚类结构。 ()
15. PCA 降维后，主成分之间是相互正交的。 ()
16. 关联规则挖掘中，支持度衡量规则的可靠性。 ()
17. 感知机无法解决异或问题。 ()
18. 神经网络中，Sigmoid 函数容易导致梯度消失。 ()
19. 批梯度下降的收敛速度通常比随机梯度下降快。 ()
20. 集成学习只能使用同一种基学习器。 ()

三、简答题（每题 6 分，共 30 分）

1. 什么是偏差和方差？它们如何影响模型的性能？
2. 简述朴素贝叶斯分类器的基本原理，并写出贝叶斯公式。
3. 解释决策树中预剪枝和后剪枝的区别及其优缺点。
4. 简述 SVM 中软间隔和硬间隔的概念及其适用场景。
5. 比较 K-means 聚类和 DBSCAN 聚类的优缺点。

四、综合题（每题 10 分，共 30 分）

1. （推导题）已知线性回归的损失函数为均方误差：

$$J(w) = \frac{1}{2m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2$$

其中假设函数为

$$h(x) = w^T x$$

请使用梯度下降法推导参数 w 的更新公式，并解释学习率 α 的作用。

2. 深度学习中的卷积神经网络（CNN）在图像处理中表现出色，请简述卷积层、池化层和全连接层的作用，并列举两个 CNN 的典型应用场景。
 3. 在模型评估中，什么是 ROC 曲线和 AUC 值？如何根据 ROC 曲线判断模型性能和根据 AUC 值判断分类模型的性能？

机器学习模拟试卷三

一、填空题（每空 1 分，共 20 分）

1. 机器学习的类型包括监督学习、无监督学习、_____学习和_____学习。
2. 模型复杂度越高，越容易导致_____；模型复杂度越低，越容易导致_____。
3. 在数据集划分中，_____集用于训练模型，_____集用于选择超参数，_____集用于最终评估。
4. 分类任务中，F1 分数是_____和_____的调和平均数。
5. 逻辑回归中，使用_____函数将线性输出映射到(0, 1)区间。
6. 朴素贝叶斯分类器中，计算后验概率时，分母 $P(X)$ 可以通过_____公式计算。
7. KNN 算法中， k 值的选择会影响模型的_____和_____。
8. 决策树 C4.5 算法使用_____作为划分标准，以克服 ID3 算法的偏好问题。
9. 随机森林中，每棵树使用_____采样和_____采样来增加多样性。
10. AdaBoost 通过调整样本权重，使得后续基分类器更关注_____的样本。
11. SVM 中，支持向量是离_____最近的样本点。
12. K-means 聚类中，簇的质心通过计算簇内样本的_____得到。
13. 层次聚类可以分为_____聚类（自下而上）和_____聚类（自上而下）。
14. PCA 中，主成分是按照_____大小排序的。
15. 关联规则中，_____衡量规则在所有交易中出现的频率，_____衡量规则的可靠性。
16. 神经网络中，反向传播算法用于计算损失函数对_____的梯度。
17. 激活函数 ReLU 的数学表达式为_____。
18. 正则化方法 Dropout 在_____阶段随机丢弃部分神经元，以减少过拟合。

二、判断题（每题 1 分，共 20 分）

(正确打“√”，错误打“×”)

1. 无监督学习不需要任何标签信息。 ()
2. 交叉验证可以完全消除过拟合。 ()
3. 逻辑回归的输出可以解释为样本属于正类的概率。 ()
4. 归一化将数据缩放到 $[0, 1]$ 区间，标准化将数据变为均值为 0、方差为 1。 ()
5. L1 正则化在损失函数中加入权重的绝对值之和。 ()
6. 朴素贝叶斯分类器可以用于多分类问题。 ()
7. KNN 算法需要对数据进行归一化处理。 ()
8. 决策树中，信息增益对取值较多的特征有偏好。 ()
9. 随机森林中，每棵决策树是相互独立的。 ()
10. AdaBoost 的基分类器必须是弱分类器。 ()
11. SVM 可以使用核函数处理非线性分类问题。 ()
12. 核函数的选择对 SVM 性能没有影响。 ()
13. 聚类分析中，所有样本必须属于某一个簇。 ()
14. DBSCAN 聚类不需要预先指定簇的数量。 ()
15. PCA 降维后，第一个主成分包含最大的方差。 ()
16. 关联规则挖掘中，置信度高的规则一定有用。 ()
17. 单层感知机可以解决非线性分类问题。 ()
18. 神经网络中，隐藏层越多，模型表达能力越强。 ()
19. 梯度下降法总能找到全局最优解。 ()
20. 集成学习中，Stacking 通过元学习器组合多个基学习器的预测结果。 ()

三、简答题（每题 6 分，共 30 分）

1. 解释什么是梯度下降法，并比较批梯度下降、随机梯度下降和小批量梯度下降的优缺点。
2. 简述决策树中 ID3、C4.5 和 CART 算法的区别。
3. 什么是集成学习？简要说明 Bagging、Boosting 和 Stacking 的基本思想。
4. 简述 K-means 聚类算法的步骤，并说明其局限性。
5. 什么是神经网络中的反向传播算法？简述其基本步骤。

四、综合题（每题 10 分，共 30 分）

1. （推导题）在神经网络中，假设使用 Sigmoid 激活函数和均方误差损失函数。请推导单层神经网络（一个隐藏层）中，损失函数对输入层到隐藏层权重 w_{ij} 的梯度公式，并说明反向传播的过程。

2. 深度学习中的循环神经网络（RNN）和长短期记忆网络（LSTM）在处理序列数据时有何优势？请简述 LSTM 如何解决 RNN 的梯度消失问题。
 3. 在模型评估中，什么是学习曲线？如何根据学习曲线判断模型是否过拟合或欠拟合？

机器学习模拟试卷四

一、填空题（每空1分，共20分）

1. 机器学习中，_____学习利用带标签的数据训练模型，_____学习利用无标签的数据发现结构。
2. 模型选择时，我们希望选择_____误差最小的模型。
3. 正则化通过向损失函数中添加_____项来防止过拟合。
4. 在分类任务中，如果正负样本数量差异很大，需要使用_____处理方法，如过采样或欠采样。
5. 逻辑回归中，代价函数通常使用_____损失函数。
6. 朴素贝叶斯分类器中，对于连续特征，通常假设其服从_____分布。
7. KNN算法中，_____距离是两个点在各个坐标轴上距离的最大值。
8. 决策树中，_____算法既可以用于分类也可以用于回归。
9. 随机森林通过_____减少模型的方差，提高泛化能力。
10. AdaBoost中，基分类器的权重根据其_____计算。
11. SVM中，_____核函数可以将数据映射到无限维空间。
12. 聚类评价指标中，_____系数衡量样本与其所属簇内其他样本的相似度。
13. DBSCAN聚类中，两个参数是扫描半径_____和最小样本数_____。
14. PCA中，每个主成分都是原始特征的_____组合。
15. 关联规则挖掘中，_____算法通过构建FP树来挖掘频繁项集。
16. 神经网络中，_____层通常用于降低特征图的维度并保留主要特征。
17. 优化算法_____结合了动量和自适应学习率的优点。
18. 在深度学习中，_____网络通过跳跃连接缓解梯度消失问题。

二、判断题（每题1分，共20分）

（正确打“√”，错误打“×”）

1. 机器学习模型在训练集上表现好，在测试集上一定表现好。 ()
2. 增加训练数据可以降低过拟合的风险。 ()
3. 逻辑回归是线性模型。 ()
4. 归一化适用于所有机器学习算法。 ()
5. L2正则化在损失函数中加入权重的平方和。 ()
6. 朴素贝叶斯分类器对缺失数据不敏感。 ()
7. KNN算法中，k值越小，模型越简单。 ()
8. 决策树剪枝可以在构建树的过程中进行，也可以在构建完成后进行。 ()
9. 随机森林中，每棵树的训练样本是通过有放回抽样得到的。 ()
10. AdaBoost中，所有基分类器的权重都相同。 ()
11. SVM的决策边界仅由支持向量决定。 ()
12. 核函数的选择依赖于领域知识。 ()
13. K-means聚类对初始质心的选择敏感。 ()
14. 层次聚类的结果可以通过树状图表示。 ()
15. PCA降维后，主成分的数量通常小于或等于特征的数量。 ()
16. 关联规则挖掘中，支持度和置信度越高，规则越有意义。 ()
17. 感知机可以使用激活函数增加非线性能力。 ()
18. 神经网络中，权重初始化为0是最佳选择。 ()
19. 随机梯度下降的收敛路径比批梯度下降更平滑。 ()
20. 集成学习中，基学习器之间的相关性越低，集成效果越好。 ()

三、简答题（每题6分，共30分）

1. 什么是模型评估中的留出法和交叉验证法？比较它们的优缺点。
 2. 解释逻辑回归中梯度下降法的参数更新过程。
 3. 简述朴素贝叶斯分类器的训练和预测步骤。
 4. 什么是决策树的过拟合？如何通过剪枝解决？
 5. 简述SVM中核技巧的基本思想，并给出两个常用核函数。

四、综合题（每题10分，共30分）

1. （推导题）在朴素贝叶斯分类器中，给定训练数据集，请推导出在条件独立性假设下，后验概率 $P(Y|X)$ 的最大化等价于最大化联合概率 $P(X, Y)$ ，并说明如何利用拉普拉斯平滑处理零概率问题。

2. 深度学习中的生成对抗网络（GAN）由哪两部分组成？简述它们的工作原理，并列举一个GAN的应用实例。
 3. 在聚类分析中，如何选择 K-means 中的 K 值？除了肘部法则，还有哪些方法可以用于确定最佳簇数？

机器学习模拟试卷五

一、填空题（每空1分，共20分）

1. 机器学习的三个基本要素是模型、_____和_____。
2. 模型评估中，_____误差是指模型在训练集上的误差，_____误差是指模型在测试集上的误差。
3. 解决过拟合的正则化方法包括L1正则化、L2正则化和_____。
4. 在二分类问题中，ROC曲线的横轴是_____，纵轴是_____。
5. 逻辑回归中，决策边界是_____函数值为0.5的线或面。
6. 朴素贝叶斯分类器中，先验概率可以通过_____估计。
7. KNN算法中，决策规则可以是多数表决或_____平均。
8. 决策树中，_____指数越小，数据集的纯度越高。
9. 随机森林中，每棵决策树在节点分裂时随机选择_____。
10. AdaBoost通过_____调整样本权重，使得后续基分类器更关注难分样本。
11. SVM中，_____距离是支持向量到决策边界距离的两倍。
12. K-means聚类中，代价函数是样本到其所属簇质心的_____之和。
13. 层次聚类中，_____方法首先将每个样本视为一个簇，然后逐步合并最近的簇。
14. PCA中，特征值表示对应主成分的_____大小。
15. 关联规则中，提升度衡量规则中项集之间的_____程度。
16. 神经网络中，_____函数可以将输出压缩到(-1,1)区间。
17. 优化算法_____使用历史梯度的平方和来调整学习率。
18. 在深度学习中，_____网络通过注意力机制处理序列数据。

二、判断题（每题1分，共20分）

（正确打“√”，错误打“×”）

1. 机器学习模型的目标是最小化训练误差。 ()
2. 交叉验证可以用于模型选择和参数调优。 ()
3. 逻辑回归可以使用梯度下降法求解。 ()
4. 数据标准化对决策树算法是必要的。 ()
5. L1正则化倾向于产生稀疏解。 ()
6. 朴素贝叶斯分类器不需要训练过程。 ()
7. KNN算法中，距离度量可以选择欧氏距离或余弦相似度。 ()
8. 决策树中，信息增益率是信息增益除以特征熵。 ()
9. 随机森林可以处理高维数据。 ()
10. AdaBoost的基分类器必须是强分类器。 ()
11. SVM只能用于分类问题。 ()
12. 核函数的选择需要根据数据和问题进行调整。 ()
13. 聚类分析中，所有聚类算法都需要预先指定簇的数量。 ()
14. DBSCAN可以发现任意形状的簇。 ()
15. PCA降维后，主成分的数量可以大于特征的数量。 ()
16. 关联规则挖掘中，频繁项集的支持度必须大于等于最小支持度阈值。 ()
17. 感知机可以用于多分类问题。 ()
18. 神经网络中，激活函数可以增加模型的非线性。 ()
19. 梯度下降法的学习率是固定的。 ()
20. 集成学习中，Boosting可以降低模型的偏差。 ()

三、简答题（每题6分，共30分）

1. 什么是偏差-方差分解？它如何帮助我们理解模型性能？
2. 简述逻辑回归的优缺点。
3. 解释KNN算法中k值的选择对模型性能的影响。
4. 比较决策树中ID3和C4.5算法的差异。
5. 简述SVM中软间隔的含义及其数学表达。

四、综合题（每题10分，共30分）

1. （推导题）在线性回归中，假设我们使用L2正则化（岭回归），损失函数为：

$$J(w) = \frac{1}{2m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n w_j^2$$

请推导出使用梯度下降法时，参数 w_j 的更新公式，并解释正则化系数 λ 的作用。

2. 请从以下算法中选择三种，比较它们的适用场景、优缺点，并说明是否需要对数据进行归一化处理：

KNN

决策树

支持向量机（线性核）

随机森林

3. 在模型评估中，什么是混淆矩阵？如何根据混淆矩阵计算准确率、查准率、查全率和F1分数？

机器学习模拟试卷六

一、填空题（每空1分，共20分）

1. 当代机器学习的重要分支包括强化学习、_____、联邦学习和图学习。
2. 强化学习中，智能体通过_____的方式在环境中学习最优策略。
3. 迁移学习利用_____的知识来辅助_____任务的学习。
4. 联邦学习的核心优势是可以在不_____的情况下进行模型训练。
5. 特征工程包括特征选择、_____和特征构建三个主要方面。
6. 在特征选择中，过滤法、_____和嵌入法是三种常用方法。
7. 优化算法Adam结合了_____和RMSProp的优点。
8. 动量法通过引入_____来加速梯度下降并减少振荡。
9. 模型解释方法中，_____方法通过构建局部线性模型来解释单个预测。
10. 异常检测中，_____算法假设正常数据服从高斯分布。
11. 推荐系统中，协同过滤可以分为基于_____的协同过滤和基于模型的协同过滤。
12. 时间序列预测中，_____模型通过差分使序列平稳。
13. GBDT使用_____作为基学习器，通过梯度提升减少残差。
14. XGBoost在GBDT基础上增加了_____项来控制模型复杂度。
15. LightGBM采用_____生长策略，相比Level-wise更高效。
16. 批量归一化（BatchNorm）通过对每层的_____进行归一化来加速训练。
17. 生成对抗网络（GAN）由生成器和_____组成，二者通过对抗训练。
18. 度量学习的目标是学习一个_____，使得相似样本距离近，不相似样本距离远。
19. 多任务学习通过共享_____来提高多个相关任务的性能。
20. 在神经网络中，_____机制使模型能够关注输入的不同部分。

二、判断题（每题1分，共20分）

（正确打“√”，错误打“×”）

1. 强化学习中，值函数表示在某个状态下执行某个动作的长期回报。 ()
2. 迁移学习要求源任务和目标任务的数据分布必须完全相同。 ()
3. 联邦学习中，所有参与方的数据都需要上传到中心服务器。 ()
4. 图神经网络可以处理非欧几里得结构的数据。 ()
5. 特征选择中的过滤法基于模型性能来选择特征。 ()
6. Adam优化算法需要手动调整学习率。 ()
7. 动量法的动量系数通常设置为0.9左右。 ()
8. SHAP方法基于博弈论，为每个特征分配重要性分数。 ()
9. 孤立森林（Isolation Forest）通过随机分割来快速识别异常点。 ()
10. 推荐系统中的冷启动问题是指新用户或新项目缺乏历史数据。 ()
11. ARIMA模型只能用于平稳时间序列的预测。 ()
12. GBDT中每棵树都试图拟合之前所有树的残差。 ()
13. XGBoost支持并行计算，因此训练速度比GBDT快很多。 ()
14. LightGBM使用直方图算法来加速特征分裂点的寻找。 ()
15. 批量归一化可以完全替代Dropout正则化。 ()
16. 变分自编码器（VAE）是一种生成模型，可以学习数据的潜在分布。 ()
17. 度量学习中，三元组损失（Triplet Loss）需要正例、负例和锚点样本。 ()
18. 多标签分类中，一个样本只能属于一个类别。 ()
19. 注意力机制最早在Transformer模型中被广泛应用。 ()
20. 模型解释性对于医疗、金融等高风险领域的应用至关重要。 ()

三、简答题（每题6分，共30分）

1. 简述强化学习的基本框架（智能体、环境、状态、动作、奖励）及其与监督学习的区别。
2. 什么是迁移学习？列举两种迁移学习的常见场景。
3. 比较GBDT、XGBoost和LightGBM的异同。
4. 解释批量归一化（BatchNorm）的作用及其在训练和测试阶段的区别。
5. 什么是注意力机制？简述其在自然语言处理中的应用。

四、综合题（每题10分，共30分）

1. （推导题）在强化学习的Q-learning算法中，Q值的更新公式为：

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$$

请解释公式中每个符号的含义，并推导出该更新公式的由来（从贝尔曼方程出发）。

2. 深度学习中的Transformer模型已经广泛应用于各个领域。请画出Transformer编码器的结构图（包含多头自注意力、前馈网络、残差连接和层归一化），并说明每个组件的作用。
 3. 假设你正在构建一个电商推荐系统，用户历史行为数据丰富，但新商品频繁上线。请设计一个混合推荐方案，解决冷启动问题，并说明使用的算法及其理由。

填空题—计算类型

1. 线性回归中，若训练集有1000个样本，每个样本有5个特征，使用最小二乘法求解时，需要计算的矩阵 $(X^T X)$ 的维度是 _____。
2. 逻辑回归使用Sigmoid函数，当 $z = w^T x + b = 2.5$ 时，预测概率 $\hat{y} =$ _____。
3. 在KNN中，若 $k = 3$ ，三个最近邻的标签分别为 $[1, 0, 1]$ ，则该样本的预测类别为 _____。
4. 决策树ID3算法中，若某特征的信息增益为0.35，条件熵为0.45，则该特征的信息熵为 _____。
5. 支持向量机中，若两个支持向量分别为 $x_1 = [1, 2]$, $x_2 = [3, 4]$ ，且类别标签分别为+1和-1，则分隔超平面的法向量 w 与这两个向量的关系是 _____。
6. K-means聚类中，若当前簇中心为 $c_1 = [2, 3]$ ，样本点 $p = [4, 7]$ ，则其到簇中心的欧氏距离为 _____。
7. PCA降维中，若某主成分的特征值为 4.5，该主成分的方差贡献率为 _____ (假设所有特征值之和为15)。
8. 朴素贝叶斯分类中，若 $P(Y = 1) = 0.3$, $P(X = 1|Y = 1) = 0.8$, $P(X = 1|Y = 0) = 0.2$ ，则后验概率 $P(Y = 1|X = 1) =$ _____。
9. 梯度下降法中，若损失函数在 $w = 2$ 处的梯度为 -1.5，学习率 $\alpha = 0.1$ ，则更新后的 $w =$ _____。
10. 正则化中，若 L2 正则化系数 $\lambda = 0.01$ ，权重向量 $w = [0.5, -0.3, 0.8]$ ，则正则化项为 _____。
11. 混淆矩阵中，若 $TP = 40$, $FP = 10$, $FN = 5$, $TN = 45$ ，则精确率 $P =$ _____，召回率 $R =$ _____。
12. 在K-means中，若数据集有500个样本，特征维度为20, $k = 5$ ，则每次迭代需要计算 _____ 次距离。

13. 逻辑回归的损失函数，若 $y = 1$, $\hat{y} = 0.9$, 则损失值为 _____。
14. 决策树CART算法中，若某节点样本类别分布为 [20,10,5], 则其基尼指数为 _____。
15. 支持向量机使用高斯核时，若 $\gamma = 0.5$, $x_i = [1, 2]$, $x_j = [3, 1]$, 则核函数值 $K(x_i, x_j)$ = _____。
16. PCA中，若原始数据维度为10，我们保留前3个主成分，则降维后的数据维度为 _____。
17. 在KNN中，若使用曼哈顿距离，两点 $x_1 = [1, 3, 5]$ 和 $x_2 = [4, 1, 2]$ 的距离为 _____。
18. 线性回归中，若使用梯度下降法，当 $m = 100$ 时，该更新称为 _____。
19. 若使用5折交叉验证，数据集有1000个样本，则每个训练子集的大小约为 _____。
20. 在集成学习中，随机森林通过 _____ 的方式降低方差。
21. 逻辑回归中，若使用L2正则化，其中 λ 称为 _____。
22. 决策树中，预剪枝是在树生成 _____ 进行剪枝。
23. 支持向量机中，软间隔允许一些样本 _____。
24. K-means聚类对初始簇中心敏感，常用的改进算法是 _____。
25. PCA中，各主成分之间是 _____ 的。
26. 朴素贝叶斯中，拉普拉斯平滑的目的是避免 _____。
27. 在模型评估中，ROC曲线下的面积称为 _____。

28. 梯度下降法中，学习率过大可能导致 _____。
29. 正则化中，L1正则化可以产生 _____ 模型。
30. 混淆矩阵中，F1分数是精确率和召回率的 _____。
31. 在KNN中，k值过大会导致 _____。
32. 决策树中，信息增益率是 _____ 与分裂信息量的比值。
33. 支持向量机中，支持向量是那些距离分隔超平面 _____ 的样本点。
34. 聚类评估中，轮廓系数越大，说明聚类效果越 _____。
35. PCA中，第一个主成分的方向是数据 _____ 最大的方向。
36. 逻辑回归中，多分类问题通常使用 _____ 策略。
37. 线性回归中，若特征之间存在多重共线性，可以使用 _____ 回归来解决。
38. 在集成学习中，AdaBoost通过调整样本权重，使得后续基学习器更关注 _____ 的样本。
39. 决策树中，后剪枝是在树生成 _____ 进行剪枝。
40. 支持向量机中，若使用线性核，则等价于 _____ 支持向量机。
41. K-means聚类中，簇内平方和 (WCSS) 的计算公式是 _____。
42. PCA中，累计方差贡献率达到 _____ 时，通常认为保留了足够的信息。
43. 朴素贝叶斯中，若特征为连续值，通常假设其服从 _____ 分布。
44. 在模型选择中，若训练集误差小，验证集误差大，说明模型出现了 _____。

45. 梯度下降法中，若使用动量（Momentum），更新公式会引入 _____。
46. 正则化中，弹性网络（Elastic Net）是 _____ 和 _____ 的正则化组合。
47. 混淆矩阵中，真正例率（TPR）等于 _____。
48. 在KNN中，若使用加权投票，权重通常与 _____ 成反比。
49. 决策树中，C4.5算法使用 _____ 作为特征选择标准。
50. 支持向量机中，若数据线性不可分，可以使用 _____。
51. 聚类算法DBSCAN将样本分为核心点、边界点和 _____。
52. PCA中，主成分是原始特征的 _____ 组合。
53. 逻辑回归中，若使用随机梯度下降（SGD），每次迭代使用 _____ 个样本。
54. 线性回归中，决定系数 R^2 的取值范围是 _____。
55. 在集成学习中，随机森林的基学习器是 _____。
56. 决策树中，若某特征有3个取值，则该特征的分裂信息量为 _____（假设样本数均等）。
57. 支持向量机中，正则化参数C越大，模型对错误分类的惩罚越 _____。
58. K-means聚类中，肘部法则（Elbow Method）用于确定 _____。
59. PCA中，若原始数据已经中心化，则协方差矩阵的计算公式为 _____。
60. 朴素贝叶斯中，若某个特征在某个类别下未出现，使用拉普拉斯平滑后，其条件概率为 _____。

61. 在模型评估中，留出法将数据集分为训练集、验证集和 _____。
62. 梯度下降法中，若损失函数是凸函数，则一定能找到 _____ 最优解。
63. 正则化中，L2正则化在损失函数中添加了权重的 _____ 项。
64. 混淆矩阵中，假正例率（FPR）等于 _____。
65. 在KNN中，若使用KD树进行加速搜索，其时间复杂度为 _____。
66. 决策树中，基尼指数越小，说明样本纯度越 _____。
67. 支持向量机中，若使用多项式核，参数 d 表示 _____。
68. 聚类评估中，调整兰德系数（ARI）的取值范围是 _____。
69. PCA中，若将数据投影到第一个主成分上，则投影后的方差为 _____。
70. 逻辑回归中，若使用批量梯度下降，每次迭代更新需要计算 _____ 样本的梯度。
71. 线性回归中，若使用正则化，目标函数由损失函数和 _____ 组成。
72. 在集成学习中，Bagging通过 _____ 减少方差。
73. 决策树中，ID3算法只能处理 _____ 特征。
74. 支持向量机中，若使用高斯核，参数 γ 越大，模型越容易 _____。
75. K-means聚类中，若某个簇中心在迭代过程中未发生变化，则算法 _____。
76. PCA中，若原始数据维度为 d ，则最多有 _____ 个非零特征值。
77. 朴素贝叶斯中，若特征之间不独立，则分类效果可能 _____。

78. 在模型选择中，交叉验证可以降低估计的 _____。
79. 梯度下降法中，若使用自适应学习率方法（如Adam），可以加快 _____。
80. 正则化中，L1正则化在损失函数中添加了权重的 _____ 项。
81. 混淆矩阵中，准确率（Accuracy）的计算公式是 _____。
82. 在KNN中，若使用切比雪夫距离，两点 $x_1 = [1, 3, 5]$ 和 $x_2 = [4, 1, 2]$ 的距离为 _____。
83. 决策树中，若使用基尼指数，某个节点的基尼指数为0，说明该节点 _____。
84. 支持向量机中，若使用软间隔，目标函数中引入了 _____ 变量。
85. 聚类算法中，层次聚类可以分为聚合式和 _____。
86. PCA中，若保留前k个主成分，则降维后的数据可以近似重构原始数据，重构误差为 _____。
87. 逻辑回归中，若使用牛顿法求解，需要计算 _____。
88. 线性回归中，若使用岭回归（Ridge Regression），其损失函数为 _____。
89. 在集成学习中，随机森林在构建每棵树时，使用的样本是通过 _____ 抽样得到的。
90. 决策树中，CART算法可以用于 _____ 和回归任务。
91. 支持向量机中，若使用SMO算法求解，每次优化 _____ 个拉格朗日乘子。
92. K-means聚类中，若初始簇中心选择不当，可能陷入 _____ 最优。
93. PCA中，若将数据投影到主成分上，投影后的特征之间 _____。

94. 朴素贝叶斯中，若特征为文本数据，通常使用 _____ 模型。
95. 在模型评估中，若正负样本比例失衡，可以使用 _____ 进行评估。
96. 梯度下降法中，若使用小批量梯度下降，批量大小（batch size）通常设置为 _____ 的倍数。
97. 正则化中，若正则化系数过大，可能导致模型 _____。
98. 混淆矩阵中，若某个类的样本数量很少，该类称为 _____。
99. 在KNN中，若使用球树（Ball Tree）进行加速搜索，其空间复杂度为 _____。
100. 决策树中，若使用代价复杂度剪枝（Cost-Complexity Pruning），其剪枝参数为 _____。

判断题

1. 线性回归中，最小二乘法求解的闭式解为 $w = (X^T X)^{-1} X^T y$ 。 (✓)
2. 逻辑回归的损失函数是凸函数，因此梯度下降一定能找到全局最优解。 (✓)
3. KNN算法在训练阶段不需要进行任何学习，只需存储训练样本。 (✓)
4. 决策树ID3算法使用信息增益率作为特征选择标准。 (✗)
5. 支持向量机中，分隔超平面由所有训练样本决定。 (✗)
6. K-means聚类算法对异常值不敏感。 (✗)
7. PCA降维后，各主成分之间的相关系数为0。 (✓)
8. 朴素贝叶斯分类器假设特征之间条件独立。 (✓)
9. 梯度下降法中，学习率越大，收敛速度越快。 (✗)
10. 正则化可以完全消除过拟合。 (✗)
11. 混淆矩阵中，精确率（Precision）和召回率（Recall）之间存在权衡关系。 (✓)
12. 在KNN中，k值越小，模型越容易过拟合。 (✓)
13. 决策树C4.5算法可以处理连续特征。 (✓)
14. 支持向量机使用核函数后，分类边界一定是非线性的。 (✗)
15. 聚类算法DBSCAN不需要预先指定簇的数量。 (✓)
16. PCA降维会损失一部分信息，因此重构数据一定与原始数据相同。 (✗)
17. 逻辑回归可以用于多分类问题，如使用Softmax回归。 (✓)
18. 线性回归中，若特征之间存在多重共线性，最小二乘法的解不稳定。 (✓)
19. 随机森林中，每棵决策树是在原始训练集的子集上训练的。 (✓)
20. AdaBoost算法中，基学习器的权重与其分类错误率成正比。 (✗)
21. 决策树中，预剪枝可能造成欠拟合。 (✓)
22. 支持向量机中，软间隔允许一些样本被错分，但惩罚项会增大。 (✓)
23. K-means聚类算法一定能收敛到全局最优解。 (✗)
24. PCA中，第一个主成分是数据方差最大的方向。 (✓)
25. 朴素贝叶斯分类器是生成模型。 (✓)

26. 梯度下降法中，随机梯度下降（SGD）的收敛速度比批量梯度下降（BGD）快。
(\times)
27. 正则化中，L1正则化倾向于产生稀疏解。 (\checkmark)
28. 混淆矩阵中，F1分数是精确率和召回率的算术平均。 (\times)
29. 在KNN中，距离度量方式（如欧氏距离、曼哈顿距离）对分类结果没有影响。 (\times)
30. 决策树中，信息增益对取值数目多的特征有偏好。 (\checkmark)
31. 支持向量机中，高斯核函数可以将数据映射到无限维空间。 (\checkmark)
32. 聚类算法中，轮廓系数越大，聚类效果越好。 (\checkmark)
33. PCA降维后，各主成分的方差之和等于原始特征的方差之和。 (\checkmark)
34. 逻辑回归的决策边界是线性的。 (\checkmark)
35. 线性回归中，若使用正则化，正则化系数越大，模型越简单。 (\checkmark)
36. 在集成学习中，Bagging可以减少偏差。 (\times)
37. 决策树中，后剪枝通常比预剪枝保留更多的分支。 (\checkmark)
38. 支持向量机中，若使用线性核，则不需要调节参数 γ 。 (\checkmark)
39. K-means聚类中，簇内平方和（WCSS）随k增大而单调递减。 (\checkmark)
40. PCA中，主成分是原始特征的线性组合，且系数向量为单位向量。 (\checkmark)
41. 朴素贝叶斯分类器需要假设特征服从某种分布（如高斯分布）。 (\checkmark)
42. 梯度下降法中，动量（Momentum）可以帮助跳出局部最小值。 (\checkmark)
43. 正则化中，L2正则化倾向于使权重接近于0但不为0。 (\checkmark)
44. 混淆矩阵中，准确率（Accuracy）对不平衡数据集不敏感。 (\times)
45. 在KNN中，若使用加权投票，距离越近的邻居权重越大。 (\checkmark)
46. 决策树中，CART算法使用基尼指数作为特征选择标准（分类任务）。 (\checkmark)
47. 支持向量机中，若数据线性可分，则硬间隔SVM的分类误差为0。 (\checkmark)
48. 聚类算法中，DBSCAN可以发现任意形状的簇。 (\checkmark)
49. PCA中，若将数据投影到前k个主成分上，则投影后的数据维度为k。 (\checkmark)
50. 逻辑回归中，Sigmoid函数的输出可以解释为概率。 (\checkmark)
51. 线性回归中，若使用梯度下降法，需要归一化特征以加速收敛。 (\checkmark)

52. 随机森林中，每棵树使用的特征子集是通过随机抽样得到的。 (✓)
53. AdaBoost算法中，每个基学习器的权重与其分类准确率成正比。 (✓)
54. 决策树中，信息增益率对取值数目少的特征有偏好。 (✓)
55. 支持向量机中，若使用多项式核，参数d越大，模型越复杂。 (✓)
56. K-means聚类中，肘部法则（Elbow Method）通过观察WCSS随k的变化来确定k。 (✓)
57. PCA中，若原始数据未中心化，则第一主成分可能指向数据的均值方向。 (✓)
58. 朴素贝叶斯中，拉普拉斯平滑可以避免因某个特征值未出现而导致的条件概率为0。 (✓)
59. 梯度下降法中，学习率衰减可以加快收敛速度。 (✓)
60. 正则化中，弹性网络（Elastic Net）结合了L1和L2正则化的优点。 (✓)
61. 混淆矩阵中，ROC曲线越靠近左上角，模型性能越好。 (✓)
62. 在KNN中，若使用KD树，搜索最近邻的时间复杂度为 $O(\log n)$ 。 (✓)
63. 决策树中，基尼指数为0表示该节点样本属于同一类别。 (✓)
64. 支持向量机中，SMO算法是一种高效求解SVM对偶问题的方法。 (✓)
65. 聚类算法中，层次聚类的距离度量方式（如单连接、全连接）会影响聚类结果。 (✓)
66. PCA中，若特征值很小，对应的主成分可以忽略。 (✓)
67. 逻辑回归中，多分类问题可以使用One-vs-Rest策略。 (✓)
68. 线性回归中，若使用Lasso回归，一些特征的系数可能变为0。 (✓)
69. 随机森林中，每棵决策树是相互独立的。 (✗)
70. AdaBoost算法对噪声数据敏感。 (✓)
71. 决策树中，剪枝可以减少过拟合。 (✓)
72. 支持向量机中，若使用高斯核，参数 γ 越小，模型越平滑。 (✓)
73. K-means聚类中，初始簇中心的选择会影响最终聚类结果。 (✓)
74. PCA中，主成分的方向是协方差矩阵的特征向量方向。 (✓)
75. 朴素贝叶斯分类器在特征相关性较强时效果可能变差。 (✓)
76. 梯度下降法中，批量梯度下降（BGD）每次迭代使用全部样本。 (✓)

77. 正则化中，正则化系数需要通过交叉验证选择。 (✓)
78. 混淆矩阵中，精确率（Precision）关注的是预测为正的样本中有多少是真正的正样本。 (✓)
79. 在KNN中，若使用余弦相似度作为距离度量，适合处理高维稀疏数据。 (✓)
80. 决策树中，C4.5算法可以处理缺失值。 (✓)
81. 支持向量机中，若使用线性核，则分类边界是线性的。 (✓)
82. 聚类算法中，DBSCAN的复杂度为 $O(n^2)$ ，不适合大规模数据。 (✓)
83. PCA中，若原始数据维度为d，则协方差矩阵是 $d \times d$ 的对称矩阵。 (✓)
84. 逻辑回归中，若使用牛顿法，需要计算海森矩阵（Hessian Matrix）。 (✓)
85. 线性回归中，若使用正则化，正则化项是权重向量的范数。 (✓)
86. 随机森林中，每棵树生长时，每个节点分裂所使用的特征子集是随机选择的。 (✓)
87. AdaBoost算法中，基学习器可以是任何分类器。 (✓)
88. 决策树中，ID3算法使用信息增益，C4.5使用信息增益率，CART使用基尼指数。
(✓)
89. 支持向量机中，若数据线性不可分，可以使用核技巧。 (✓)
90. K-means聚类中，簇中心更新为该簇所有样本的均值。 (✓)
91. PCA中，若将数据投影到主成分上，投影后的数据保留了原始数据的大部分信息。
(✓)
92. 朴素贝叶斯中，若特征为连续值，可以使用高斯朴素贝叶斯。 (✓)
93. 梯度下降法中，随机梯度下降（SGD）的更新方向方差较大。 (✓)
94. 正则化中，L1正则化可以用于特征选择。 (✓)
95. 混淆矩阵中，召回率（Recall）关注的是实际为正的样本中有多少被预测为正。 (✓)
96. 在KNN中，若使用马氏距离，考虑了特征之间的相关性。 (✓)
97. 决策树中，过拟合的决策树往往深度较大。 (✓)
98. 支持向量机中，若使用软间隔，正则化参数C控制对错分样本的惩罚程度。 (✓)
99. 聚类算法中，K-means的簇形状是球形的。 (✓)
100. PCA中，若保留所有主成分，则重构数据与原始数据相同。 (✓)

机器学习总复习—2025-12-25版本

一、填空题（每空1分，共20分）

1. 机器学习是利用_____，通过计算技术模拟人类学习过程的学科。
2. 监督学习中，若标签是离散值，则该任务属于_____任务。
3. 过拟合是指模型在_____集上表现良好，但在_____集上表现较差。
4. 线性回归中，最小化均方误差的方法之一是_____。
5. 梯度下降的三种形式分别是：批量梯度下降、_____、小批量梯度下降。
6. 朴素贝叶斯基于_____假设，即各特征在给定类别下条件独立。
7. KNN算法的三个基本要素是：_____、距离度量和决策规则。
8. 决策树中，ID3算法使用_____作为特征选择准则。
9. 随机森林是_____（Bagging / Boosting）方法的一种。
10. SVM中，使用_____核函数可将数据映射到高维空间。
11. K均值聚类中，肘部法则用于确定_____。
12. PCA降维的核心是保留数据中_____最大的方向。
13. 深度学习中的前向传播是指信息从_____层向_____层传递的过程。
14. 感知机是单层神经网络，只能处理_____问题。
15. 反向传播算法利用_____法则更新网络权重。
16. Transformer模型的核心机制是_____。
17. 欠拟合的解决方法之一是_____模型复杂度。

二、判断题（每题1分，共20分）

1. 机器学习的目标是让模型在训练集上达到100%准确率。 ()
2. 监督学习需要标签，而无监督学习不需要。 ()
3. 过拟合可以通过增加训练数据来缓解。 ()
4. 线性回归可以用于分类问题。 ()
5. 梯度下降法一定能找到全局最优解。 ()
6. 朴素贝叶斯属于生成模型。 ()
7. KNN算法对异常值不敏感。 ()
8. 决策树的剪枝是为了减少过拟合。 ()
9. 随机森林中的每棵树是独立训练的。 ()
10. AdaBoost中，分类错误的样本权重会降低。 ()
11. SVM只能处理线性可分数据。 ()
12. 核函数可以隐式地将数据映射到高维空间。 ()
13. K均值聚类需要预先指定聚类数量。 ()
14. PCA降维后数据维度越低越好。 ()
15. 深度学习是机器学习的一个分支。 ()
16. 感知机可以解决异或问题。 ()
17. 卷积神经网络适用于图像处理任务。 ()
18. 注意力机制可以增强模型对关键信息的关注。 ()
19. 反向传播只能用于全连接网络。 ()
20. 集成学习中，Bagging主要降低方差，Boosting主要降低偏差。 ()

三、简答题（每题6分，共30分）

1. 简述深度学习与机器学习的联系与区别。
2. 说明KNN算法与K均值聚类的主要区别，并简述KNN算法的求解步骤、K均值聚类的求解步骤。
3. 简述集成学习中Bagging、Boosting和Stacking的联系与区别。
4. 简述决策树的原理、构造过程及其优缺点。
5. 请说明最小二乘法和梯度下降法在线性回归中的区别与联系。

四、综合题（共30分）

1. 支持向量机推导题（10分）

给定线性可分数据集，支持向量机的目标是找到一个超平面，使得两类样本之间的间隔最大。请推导SVM的优化目标函数，并解释“支持向量”的含义。

对于线性可分的二分类问题，支持向量机的目标是最大化间隔：

- (1) 写出间隔的数学定义；
- (2) 推导出SVM的原始优化问题；
- (3) 解释什么是支持向量，并说明其在模型中的作用。

2. 降维算法计算题 (10分)

给定数据矩阵

$$X = \begin{pmatrix} 1 & 2 & 2 \\ 4 & 4 & 5 \end{pmatrix}$$

1. 使用SVD对其进行分解，写出 U 、 Σ 、 V^T ；
2. 使用PCA将其降至1维，写出降维后的数据。

已知数据矩阵：

$$X = \begin{pmatrix} 2 & 4 \\ 4 & 6 \end{pmatrix}$$

其中每一行表示一个样本，每一列表示一个特征。

1. 对矩阵 X 进行奇异值分解 (SVD)，写出其分解形式

$$X = U\Sigma V^T,$$

并给出矩阵 U 、 Σ 和 V^T ；

2. 先对数据矩阵 X 进行去均值（中心化）处理，然后基于中心化后的数据使用主成分分析 (PCA) 方法将数据降至一维，并写出降维后的结果。

3. 机器学习应用设计题（10分）

假设你要构建一个“垃圾邮件分类系统(分类)”，请设计一个完整的机器学习流程，包括：

- 数据收集与预处理
- 特征提取方法
- 模型选择与训练
- 模型评估指标
- 可能的优化策略

其他场景：房价预测系统(回归)、课堂点名人脸检测(检测)、图书馆人流量预测(预测)

一、填空题（补充）

1. 机器学习中，_____误差指模型在训练集上的误差，_____误差指模型在新样本上的误差。
2. 若任务标签为连续值，则该监督学习任务为_____。
3. 解决过拟合的常用方法有：增加数据、_____、正则化等。
4. 线性回归中，使用最小二乘法求解时，需要计算矩阵_____（写出具体形式）。
5. 随机梯度下降每次更新使用_____个样本。
6. 朴素贝叶斯中，为防止概率为0，可使用_____平滑方法。
7. KNN算法中， k 值过小容易导致_____， k 值过大容易导致_____。
8. 决策树C4.5算法使用_____作为特征选择准则，以克服ID3算法的偏好。
9. 随机森林通过_____方式生成多个训练子集。
10. SVM中，高斯核函数的参数 γ 越大，模型的拟合能力越_____。
11. 聚类分析中，DBSCAN是基于_____的聚类算法。
12. PCA降维后，各主成分之间是_____的。
13. 深度学习模型中，Dropout是一种_____技术，用于防止过拟合。
14. 卷积神经网络中，卷积操作主要用于提取_____特征。
15. 注意力机制的核心是计算_____权重，以聚焦重要信息。

二、判断题（补充）

1. 无监督学习就是完全没有标签的学习。 ()
2. 欠拟合可以通过减少模型复杂度来解决。 ()
3. 正则化项在损失函数中用于控制模型复杂度。 ()
4. 逻辑回归是一种分类算法，其输出是概率值。 ()
5. 梯度下降法一定比最小二乘法更优。 ()
6. 贝叶斯分类器要求特征之间相互独立。 ()
7. KNN算法在训练阶段不需要学习模型参数。 ()
8. 决策树容易过拟合，因此不需要剪枝。 ()
9. Bagging和Boosting都可以并行训练基学习器。 ()
10. SVM只能用于二分类问题。 ()
11. 核函数的选择对SVM性能没有影响。 ()
12. K均值聚类对初始聚类中心的选择不敏感。 ()
13. 降维一定会损失信息。 ()
14. 深度学习模型层数越多越好。 ()
15. 反向传播算法只能用于全连接神经网络。 ()

深度学习

一、填空题

1. 机器学习的三要素是模型、学习准则和 _____。
2. 为了避免过拟合，常用的正则化方法有L1正则化、L2正则化和 _____。
3. 在Logistic回归中，常用的损失函数是 _____。
4. 前馈神经网络中，信号从输入层向输出层 _____ 传播。
5. 通用近似定理指出，一个具有至少一个隐藏层的前馈神经网络可以近似任何 _____ 函数。
6. 反向传播算法是 _____ 的一种特殊形式。
7. 卷积神经网络的三个结构特性是局部连接、权重共享和 _____。
8. 在CNN中，用于降维的操作是 _____。
9. ResNet通过引入 _____ 结构缓解了深层网络的训练难题。
10. RNN相比前馈网络的优势是可以处理 _____ 数据。
11. LSTM通过 _____ 机制缓解了长期依赖问题。
12. 梯度消失问题常出现在使用 _____ 激活函数的深层网络中。
13. Adam优化器结合了 _____ 和RMSprop的优点。
14. Dropout在训练时随机丢弃一部分神经元，可以看作是一种 _____ 方法。
15. 自注意力机制中，每个位置可以关注到输入序列的 _____ 位置。
16. 变分自编码器（VAE）是一种 _____ 生成模型。
17. GAN由生成器和 _____ 组成，两者进行对抗训练。
18. 强化学习中，智能体通过与环境交互来最大化 _____。
19. 马尔可夫决策过程（MDP）包含状态、动作、_____ 和奖励。
20. 深度Q网络（DQN）使用 _____ 网络来近似Q值函数。
21. 在概率图模型中，贝叶斯网络属于 _____ 图模型。
22. 玻尔兹曼机是一种 _____ 图模型。
23. 受限玻尔兹曼机（RBM）的结构是 _____ 图。

24. 期望最大化 (EM) 算法包含E步和 _____ 步。
25. 支持向量机 (SVM) 试图找到一个 _____ 来分隔两类数据。
26. 感知器使用 _____ 函数作为激活函数。
27. Softmax函数将输出转换为 _____ 分布。
28. 在机器学习中, _____ 误差衡量模型在训练集上的表现。
29. _____ 误差衡量模型在未知数据上的表现。
30. 当模型过于简单时, 容易出现 _____ 拟合。
31. 当模型过于复杂时, 容易出现 _____ 拟合。
32. 集成学习通过组合多个模型来降低 _____。
33. 在神经网络中, _____ 用于控制神经元输出的非线性变换。
34. 反向传播中, 梯度通过 _____ 法则逐层传递。
35. 池化操作通常包括最大池化和 _____ 池化。
36. 空洞卷积通过插入 _____ 来增大感受野。
37. 文本卷积模型中, 滤波器在 _____ 维度上进行滑动。
38. 循环神经网络中, 隐藏状态通过 _____ 传递。
39. 双向RNN同时考虑过去和 _____ 的信息。
40. 注意力机制中, _____ 函数用于计算查询与键的相似度。
41. Transformer模型完全基于 _____ 机制。
42. 神经图灵机包含控制器和外部 _____。
43. 自编码器的目标是最小化 _____ 误差。
44. GAN的训练是一个 _____ 博弈过程。
45. 在VAE中, 隐变量通常服从 _____ 分布。
46. 强化学习中, _____ 用于衡量未来奖励的当前价值。
47. 策略梯度方法直接优化 _____ 函数。
48. 在CNN中, 步长 (stride) 控制滤波器的 _____。
49. 批归一化 (BatchNorm) 用于缓解 _____ 偏移问题。
50. 学习率衰减策略包括步长衰减、 _____ 衰减等。

51. 在机器学习中，当训练误差很小而测试误差很大时，我们称模型出现了 _____ 问题。
52. 为了避免过拟合，我们可以使用 _____ 正则化，它通过向损失函数中添加权重绝对值和来实现。
53. 在深度学习中， _____ 算法通过反向传播误差并逐层更新权重，以最小化损失函数。
54. 卷积神经网络中的 _____ 层用于减少特征图的空间尺寸，常见的有最大池化和平均池化。
55. 循环神经网络在处理长序列时容易遇到 _____ 问题，导致难以捕捉长期依赖。
56. LSTM通过引入 _____、遗忘门和输出门来控制信息的流动。
57. 在Transformer模型中， _____ 机制允许模型在处理每个词时关注输入序列中的所有词。
58. 生成对抗网络（GAN）由生成器和 _____ 组成，两者在训练过程中相互对抗。
59. 变分自编码器（VAE）在训练时除了重构损失外，还引入了 _____ 损失，使得隐变量接近先验分布。
60. 在强化学习中， _____ 是指智能体在某一状态下选择动作的概率分布。
61. 深度Q网络（DQN）使用 _____ 网络来稳定训练，通过延迟更新目标Q值。
62. 批量归一化（Batch Normalization）通过对每个批次的输入进行 _____ 和缩放，加速网络训练。
63. Dropout在训练阶段随机丢弃一部分神经元，相当于进行了 _____ 学习。
64. 残差网络（ResNet）通过引入 _____ 连接，使得深层网络更容易优化。
65. 在注意力机制中， _____ 函数用于计算查询向量与键向量之间的相关性。
66. 自注意力（Self-Attention）中，每个位置的输出是值向量的加权和，权重由查询与键的 _____ 决定。
67. 神经图灵机（Neural Turing Machine）包含控制器和外部 _____，可以通过注意力机制进行读写。
68. 受限玻尔兹曼机（RBM）是一种 _____ 图模型，由可见层和隐藏层组成。
69. 在概率图模型中， _____ 网络是有向无环图，而马尔可夫随机场是无向图。

70. 在深度学习中，_____ 优化器结合了动量法和RMSprop的优点，通常表现优异。
71. 全连接层中，每个神经元与前一层的所有神经元 _____。
72. 在神经网络中，全连接层的参数数量主要取决于 _____ 和输出维度。
73. 感知机是一种最基本的神经网络模型，本质上是一个 _____ 分类器。
74. 单层感知机只能解决 _____ 的分类问题。
75. 卷积神经网络通过 _____ 连接和权重共享来减少参数数量。
76. 在CNN中，池化层的主要作用是降低特征图的 _____。
77. 注意力机制通过计算 _____ 与键之间的相关性来分配权重。
78. 自注意力机制中，序列中每个位置都可以关注到 _____ 的位置。
79. Transformer模型完全基于 _____ 机制，而不依赖循环结构。
80. 在Transformer中，多头注意力可以从不同的 _____ 学习特征表示。
81. 向前传播过程中，数据从输入层经过若干隐藏层最终到达 _____。
82. 在向前传播中，每一层的输出通常由线性变换和 _____ 共同决定。
83. 向后传播算法利用 _____ 法则计算损失函数对各层参数的梯度。
84. 在向后传播过程中，梯度从 _____ 层逐层向前传播。

二、判断题

1. 机器学习就是优化问题。 ()
2. 过拟合是指模型在训练集上表现差，在测试集上表现好。 ()
3. L1正则化会使部分权重变为0。 ()
4. 感知器可以解决非线性可分问题。 ()
5. Softmax回归只能用于二分类问题。 ()
6. 前馈神经网络中可以有循环连接。 ()
7. 反向传播算法只能用于全连接网络。 ()
8. 卷积神经网络的全连接层参数通常比卷积层多。 ()
9. 池化层可以减少特征图的尺寸。 ()
10. ResNet的残差连接缓解了梯度消失问题。 ()
11. RNN可以处理变长序列数据。 ()
12. LSTM比普通RNN更容易出现梯度爆炸。 ()
13. Adam优化器不需要手动调整学习率。 ()
14. Dropout在测试时也会随机丢弃神经元。 ()
15. 自注意力机制的计算复杂度与序列长度平方成正比。 ()
16. 变分自编码器（VAE）是一种判别模型。 ()
17. GAN的生成器和判别器是同时训练到最优的。 ()
18. 强化学习的目标是最小化累计奖励。 ()
19. 马尔可夫决策过程中，下一个状态只与当前状态有关。 ()
20. DQN使用目标网络来稳定训练。 ()
21. 贝叶斯网络是无向图模型。 ()
22. 玻尔兹曼机是全连接的概率图模型。 ()
23. EM算法保证收敛到全局最优解。 ()
24. 支持向量机只能处理线性可分数据。 ()
25. 特征工程在深度学习中不再重要。 ()

26. ReLU激活函数可以缓解梯度消失问题。 ()
27. 交叉熵损失函数常用于回归问题。 ()
28. 批归一化可以加速训练并提高泛化能力。 ()
29. 学习率越大，模型训练越快，效果越好。 ()
30. 早停 (Early Stopping) 是一种正则化方法。 ()
31. 集成学习只能用于分类任务。 ()
32. 神经网络可以拟合任意复杂的函数。 ()
33. 梯度下降法一定能找到全局最优解。 ()
34. 卷积核的大小影响感受野的大小。 ()
35. 空洞卷积会增加参数数量。 ()
36. RNN在时间维度上是共享参数的。 ()
37. 注意力机制可以用于图像分类任务。 ()
38. 神经图灵机是不可微分的。 ()
39. 自编码器通常用于降维和去噪。 ()
40. GAN的训练过程非常稳定。 ()
41. VAE的隐变量是离散的。 ()
42. 强化学习不需要标签数据。 ()
43. 策略梯度方法是基于值函数的方法。 ()
44. 深度学习模型越复杂，泛化能力一定越好。 ()
45. 数据增强可以提高模型的鲁棒性。 ()
46. 权重衰减等价于L2正则化。 ()
47. 学习率调度是超参数优化的一部分。 ()
48. 神经网络的所有参数都需要初始化。 ()
49. 卷积神经网络只能处理图像数据。 ()
50. 循环神经网络可以用于时间序列预测。 ()