1. **机器学习概念：**机器学习是一门多领域交叉学科，它专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。主要类型有监督学习，无监督学习，强化学习。
2. **评价指标：**

（1）分类任务评价指标：准确率，精确率，召回率，F1-Score

1. 回归任务评价指标：均方误差，均方根误差，平均绝对误差

**机器学习的定义是什么？常见的机器学习的类别有哪些？谈谈你对机器学习的理解。**

机器学习是人工智能的一个分支，它专注于开发算法和统计模型，使得计算机系统能够从数据中自动学习并改进经验，而无需明确编程指令。常见的机器学习类别包括监督学习、无监督学习、半监督学习和强化学习。监督学习通过标记的数据集来训练模型预测输出；无监督学习旨在从未标记的数据中找出隐藏的模式或内在结构；半监督学习结合了少量标记数据与大量未标记数据进行训练；强化学习则是通过奖励机制让模型在环境中学习最优行为策略。我认为机器学习是连接数据与智能决策的桥梁，它不仅推动了科技的发展，也在不断改变着我们解决问题的方式和对世界的理解。

**输入空间，输出空间和假设空间之间的关系是怎么样的？（简答，只用一段话回答）**

输入空间、输出空间和假设空间是机器学习中的三个核心概念，其中输入空间是指所有可能的输入数据点的集合，输出空间是模型预测结果的所有可能取值的集合，而假设空间则是指所有可能的模型或函数的集合，这些模型将输入映射到输出。三者之间的关系在于：通过选择一个特定的算法和设定其参数，我们在假设空间中确定了一个具体的映射规则，用于将输入空间中的元素转换为输出空间中的对应元素。换句话说，机器学习的过程就是在给定的输入输出对的基础上，在假设空间中搜索最佳的映射函数。

**简述机器学习三要素**

机器学习的三要素是指模型、策略和数据。模型指的是用来描述输入数据与输出结果之间关系的数学结构；策略涉及定义一个目标函数（如损失或成本函数）以及优化算法，用于评估和调整模型参数以最小化预测误差；数据则是训练模型的基础资源，通过大量的样本数据让模型学习到有效的映射规则。这三个要素相互作用，共同决定了机器学习系统的性能和效果。

**常见损失函数有哪些？各有什么特点。**

常见的损失函数包括均方误差（MSE），常用于回归问题，它衡量预测值与真实值之间的平方差的平均值，对异常点敏感；交叉熵损失，适用于分类任务，能够有效度量概率分布间的差异，促进模型学习到更清晰的类别边界；以及绝对误差（MAE），作为MSE的替代，它计算预测值和实际值之间绝对差值的平均，对异常值的鲁棒性更强。此外，还有合页损失（Hinge Loss）用于支持向量机，鼓励决策边界上的样本具有更大的间隔；对数损失（Log Loss）则是在概率估计中确保预测概率尽可能接近真实标签。每个损失函数都有其适用场景，选择合适的损失函数对于模型性能至关重要。

**模型的方差和偏差是如何定义的？模型的期望泛化误差与哪些因素有关，曲线图是怎样的？**

模型的方差衡量的是模型预测值的变化程度，即模型对训练数据的小波动有多敏感，高方差可能导致过拟合；偏差则是指模型预测的平均值与真实值之间的系统性差异，高偏差通常意味着模型欠拟合。模型的期望泛化误差与偏差、方差以及不可约误差（噪声）有关，随着模型复杂度增加，偏差减小而方差增大，形成U型曲线，理想模型位于偏差和方差的平衡点，此时泛化误差最小。

**简述梯度下降算法的原理，并简要叙述梯度下降算法的影响因素。**

梯度下降算法通过迭代更新模型参数来最小化损失函数，其原理是沿着损失函数梯度的反方向逐步调整参数值，因为该方向上损失减少最快。影响梯度下降效果的因素包括学习率——决定了每次迭代中参数更新的步伐大小，过大可能导致振荡不收敛，过小则收敛缓慢；初始参数值——不同的起点可能找到不同的局部最小值；以及损失函数的形状——存在多个局部极小值时，算法可能陷入非全局最优解。此外，数据规模和特征缩放也会影响梯度下降的速度和稳定性。

**分析批梯度下降法、随机梯度下降法和mini-batch梯度下降法的优缺点。**

批梯度下降法使用全部数据来计算每次迭代的梯度，其优点是收敛稳定且能到达全局最小值（对于凸问题），但缺点是当数据集很大时计算成本高。随机梯度下降法每次仅用一个样本来更新参数，这使得它在大规模数据集上更快，并且可以跳出局部最小值，然而它的路径较为波动，可能需要更多迭代才能收敛。Mini-batch梯度下降法则结合了两者的优点，在每次迭代中使用小批量样本，既能减少更新的波动性又保持较快的计算速度，因此它是实践中最常用的方法。

**感知机是一个线性分类器，是否可以使用感知机来解决非线性问题？**

感知机作为一种线性分类器，其决策边界是线性的，因此它不能直接处理非线性可分问题。然而，通过引入核技巧或增加特征空间的维度（例如使用多项式特征），可以将原始的非线性问题转换为高维空间中的线性问题，从而使得感知机或其变体能够间接解决非线性分类任务。在实际应用中，更常采用如支持向量机结合核函数等方法来应对非线性问题。

**针对几何间隔和函数间隔，硬间隔和软间隔，谈谈你的理解。**

几何间隔指的是样本点到分类超平面的垂直距离，它衡量了分类的确信程度；函数间隔则是几何间隔与单位向量的乘积，反映了模型预测输出和真实标签之间的契合度。硬间隔支持向量机力求所有数据点都严格正确分类，即间隔边界外无任何数据点，适用于线性可分的情况；而软间隔支持向量机允许部分数据点位于间隔之内甚至被误分类，通过引入松弛变量来处理噪声和非线性情况，提高了模型的泛化能力。软间隔通过平衡分类准确性与最大间隔的原则，更适应现实中的复杂数据分布。

**SVM是如何解决线性不可分问题的？请以异或函数为例来阐明你的结论。**

支持向量机（SVM）通过使用核技巧解决线性不可分问题，例如对于异或函数这样的非线性分类问题，原始二维空间中无法找到一个线性超平面将不同类别的点分开。SVM利用核函数（如多项式核或径向基函数核RBF）将输入数据映射到高维特征空间，在这个新的空间中原本线性不可分的数据变得线性可分，然后在高维空间中构建最优分类超平面。对于异或问题，这种转换可以成功地实现数据点的正确分类，即使在原始空间中这些点是交错分布的。

**线性回归中的岭回归模型应用在什么场合？正则化参数对模型有何影响？**

岭回归模型（Ridge Regression）适用于线性回归中存在多重共线性问题或特征数量较多的场合，它通过在损失函数中添加L2正则化项来惩罚过大的回归系数，从而防止模型过拟合。正则化参数（通常标记为λ或α）控制着正则化强度：当该参数值较小时，模型接近普通最小二乘法，可能会有过拟合的风险；而当参数值较大时，回归系数会被压缩得更小，甚至趋近于零，这有助于减少模型复杂度和提高泛化能力，但若过大则可能导致欠拟合。因此，选择合适的正则化参数对于平衡偏差与方差、优化模型性能至关重要。

**Softmax回归模型用在什么场合？Softmax回归模型一定比多个逻辑回归模型更合适么？为什么？**

Softmax回归模型主要用于多分类问题的场合，它能够将输入映射到多个类别的概率分布上，并选出具有最高概率的类别作为预测结果。相较于构建多个二分类逻辑回归模型，Softmax回归不一定总是更合适；虽然它可以更优雅地处理多分类任务并保证输出概率和为1，但在某些情况下，如果类别之间本质上是二元对立或多组二元分类问题，则分别训练多个逻辑回归模型可能更简单且效果相当。选择哪种模型应根据具体应用场景、数据特性和实验结果来决定。

**对于多分类问题，Softmax回归模型是否一定比Logistic回归模型（分解成多个二分类任务）更合适？**

对于多分类问题，Softmax回归模型并不一定比将Logistic回归模型分解成多个二分类任务更合适。Softmax回归能够同时处理多个类别并确保输出概率的和为1，适合类别间互斥的情况；而通过一对其余（One-vs-Rest, OvR）或多对多（One-vs-One, OvO）策略分解成多个二分类任务的Logistic回归，在某些场景下可能更简单易行，尤其是当类别之间存在层次结构或非互斥关系时。因此，哪种方法更合适取决于具体的应用场景和数据特性。

**简述SVM与Logistic回归模型的区别，并其中最本质的区别。**

SVM（支持向量机）和Logistic回归模型的主要区别在于它们的目标和输出：SVM旨在找到最大间隔超平面来分离不同类别的数据，专注于决策边界的位置，并输出类别标签；而Logistic回归则估计样本属于某一类别的概率，通过逻辑函数将线性组合的输入映射到(0,1)区间。最本质的区别是，SVM是一个侧重于边界最大化且输出离散标签的判别模型，而Logistic回归是一个基于概率的模型，更关注于预测事件发生的概率。

**深度神经网络中常见的激活函数有哪些，各有什么优缺点？**

深度神经网络中常见的激活函数包括Sigmoid、Tanh、ReLU及其变体（如Leaky ReLU, PReLU和ELU）以及Softmax。Sigmoid将值压缩到(0,1)之间，适用于二分类问题的输出层，但容易导致梯度消失；Tanh是Sigmoid的零中心版本，有助于加速收敛，同样存在梯度消失的问题。ReLU通过将所有负值设为零来引入非线性，计算简单且能有效缓解梯度消失，但可能导致“死亡神经元”现象；其变体则试图解决ReLU的这一缺陷。Softmax通常用于多分类问题的输出层，可以将预测值转换成概率分布。选择激活函数时需考虑网络的深度、任务类型及优化挑战等因素。

**梯度消失问题产生的原因是什么？如何应对深度神经网络中的梯度消失问题？**

梯度消失问题产生的原因是，在深度神经网络中，当使用如Sigmoid或Tanh这样的激活函数时，如果权重更新通过链式法则回传到较早层，梯度会变得非常小，导致这些层的参数几乎不更新。为应对这一问题，可以采用ReLU及其变体等激活函数，它们对于正输入有恒定的梯度；初始化权重时使用如Xavier或He初始化方法；应用批归一化（Batch Normalization）技术稳定和加速训练过程；或者设计残差网络（ResNets）等架构引入捷径连接（skip connections），帮助梯度更有效地传递。

**观看教学视频，阅读相关资料，进行反向传播算法的推导。并以此为基础，简述反向传播算法。**

反向传播算法是训练神经网络的核心方法，它基于链式法则计算损失函数相对于每个权重的梯度。该算法首先进行一次前向传播以计算预测输出和损失；然后从输出层开始，逐层向后计算并传播误差梯度，同时根据计算出的梯度调整权重以最小化损失。通过迭代这个过程，模型能够学习到数据中的模式。简而言之，反向传播使得神经网络可以通过不断更新权重来优化性能，其中梯度下降法或其变体用于指导权重的更新方向和幅度。

**简述进行概率模型参数估计的统计学界不同学派的观点。**

在概率模型参数估计中，频率学派和贝叶斯学派持有不同的观点：频率学派认为参数是固定的未知常数，通过数据来估计这些参数的最优值，常用方法如最大似然估计；而贝叶斯学派则视参数为随机变量，结合先验分布和观测数据来计算参数的后验分布，体现了对参数不确定性的量化。两大学派的方法反映了不同哲学理念下的统计推断途径。

**决策树如何选择最优划分属性？请结合ID3算法、C4.5算法或CART算法加以说明。**

决策树选择最优划分属性时，ID3算法使用信息增益来评估每个属性，即通过计算划分前后数据集纯度的提升程度，选择信息增益最大的属性进行分裂；C4.5算法在ID3基础上改进，采用信息增益比以减少对多值属性的偏好；而CART（分类与回归树）算法则根据基尼指数或均方误差选择最优划分，旨在最小化子节点的不纯度或误差。这三种算法的核心思想都是寻找能够最有效地提高模型纯度或预测准确性的属性。

**简要叙述下集成学习中的Bagging与Boosting的算法思想。与Bagging集成策略相比，Boosting集成策略有哪些优点和缺陷？**

Bagging通过 bootstrap 方法随机采样生成多个训练集，构建若干个并行的基学习器（如决策树），最终预测结果通过投票或平均得出，以降低方差和防止过拟合；而Boosting采用串行方式，每个学习器基于前一个学习器的错误进行训练，不断调整样本权重使后续模型更关注误分类样本，以此减少偏差。相较于Bagging，Boosting能够更好地适应数据中的复杂模式，但在面对噪声数据时更容易过拟合，且由于其串行训练机制，计算成本较高、训练时间较长。

**名词解释：信息熵，拉普拉斯修正（Laplacian Correction）**

信息熵是衡量一组数据不确定性或混乱程度的指标，定义为所有可能事件发生概率的负加权和的对数，用于量化系统的平均信息量；拉普拉斯修正是在计算概率时为了避免零频率问题（即某些类别未在训练集中出现而导致的概率估计为零），对每个类别的计数加上一个常数（通常是1），同时增加分母以保持概率和为1，确保模型对未观察到的数据也有一定的鲁棒性。

**简述流形学习中测地线距离的概念及计算方法。**

在流形学习中，测地线距离是指数据点沿着数据流形表面的最短路径长度，而非欧氏空间中的直线距离，它能更好地反映高维数据内在的低维结构。计算测地线距离的方法之一是首先构建数据点之间的邻接图（如k近邻图），然后通过图上的最短路径算法（例如Dijkstra算法或Floyd-Warshall算法）来估算点与点间的测地线距离。

**分析L1和L2范数正则化的优缺点。**

L1范数正则化能产生稀疏解，使模型中不重要的特征权重变为零，有助于特征选择和提高模型的可解释性，但可能导致估计偏差增大；L2范数正则化通过平方项惩罚较大权重，可以防止任何单个权重过大，从而减少模型过拟合，不过它不会产生稀疏性，所有特征都会保留一定的权重。因此，L1有助于简化模型，而L2则更侧重于稳定性和整体性能提升。

**为什么要进行特征选择？特征选择主要有哪3种方式？3种特征选择方式的主要区别是什么？**

（1）去除不相关特征降低学习难度；减轻维度灾难（仅在少量属性上构建模型、增加有限样本在属性空间的分布密度）；（2）过滤式、包裹式、嵌入式；（3）过滤式：先对数据集进行特征选择，其过程与后续学习无关；包裹式：就是一个分类器，它是将后续的学习器的性能作为特征子集的评价标准；嵌入式：是学习器自主选择特征。

**朴素贝叶斯分类器采用了什么假设简化建模过程？引入该假设主要解决的困难是什么？若某个属性值在训练集中没有与某个类同时出现过，则直接计算该类概率会出现问题，解决的策略是什么？**

（1）属性条件独立性假设；（2）类条件概率是所有属性上的联合概率，难以从有限的训练样本估计获得；（3）拉普拉斯修正。