Q1 请用通俗语言解释为什么机器学习中要对数据做归一化或标准化处理

在机器学习中，我们通常需要处理不同类型、不同尺度的特征数据。如果不对这些数据进行归一化或标准化处理，会出现以下问题：

数据间的量级不同，会对模型的训练产生影响。比如，某个特征的取值范围远大于其他特征的取值范围时，模型在训练时可能会更加关注该特征，从而导致其他特征的权重被低估，影响模型的准确性。

数据的分布不同，会对模型的训练产生影响。比如，某个特征的数据分布呈现正态分布，而其他特征的数据分布不符合正态分布时，模型的性能可能会受到影响，从而导致模型的泛化能力下降。

因此，为了解决这些问题，我们需要对数据进行归一化或标准化处理，使得所有特征数据的取值范围和分布都大致相同。这样，在模型训练时，各个特征的权重才能够更加均衡地影响模型的输出，从而提高模型的准确性和泛化能力。

Q2 在机器学习中， z-score变换后的数据一定是正态分布吗？

在机器学习中，将数据进行z-score标准化后，其分布形态会发生变化，但并不一定会变成正态分布。

Z-score标准化是一种常用的数据预处理技术，它通过将原始数据转化为均值为0，标准差为1的分布来消除不同特征值的量纲影响，使得不同特征之间具有可比性。但是，它并不会改变原始数据的分布形态，只是对原始数据进行了线性变换，因此如果原始数据不是正态分布的，那么经过z-score标准化后也不会是正态分布的。

实际上，只有当原始数据本身是正态分布时，z-score标准化后的数据才是正态分布的。对于非正态分布的数据，可以考虑使用其他的变换方法来使其更接近于正态分布，例如对数变换、Box-Cox变换等。

Q3 在线性回归中，数据进行z-score标准化后，梯度下降的等高线形状会变得近似圆形，为什么？

在线性回归中，将数据进行z-score标准化后，梯度下降的等高线形状会变得近似圆形，这是因为标准化后的数据使得不同特征的尺度大小变得相似，减小了等高线的椭圆度，使其更加接近于圆形。

在线性回归中，梯度下降是一种常用的优化方法，其目的是最小化损失函数。而损失函数的等高线表示的是在参数空间中，所有可以达到相同损失值的参数组成的等高线。在标准

化前的数据中，不同特征之间的尺度大小差异较大，导致等高线的形状较扁，这意味着优化算法在不同特征方向上的收敛速度差异较大，可能会出现优化过程中步长不稳定、收敛速度慢等问题。

而在标准化后的数据中，不同特征之间的尺度大小变得相似，等高线的形状更接近于圆形，使得优化算法在不同特征方向上的收敛速度更加一致，可以更加稳定地收敛到最优解。因此，在线性回归中，进行z-score标准化后可以提高模型的稳定性和收敛速度，避免优化过程中的问题。

Q4 请总结机器学习中 随机梯度下降和批量梯度下降的适用场景与性能差异

随机梯度下降（Stochastic Gradient Descent，SGD）和批量梯度下降（Batch Gradient Descent，BGD）是机器学习中两种常用的优化算法。它们的适用场景和性能差异如下：

适用场景:

SGD适用于大规模数据集和高维特征空间，因为SGD每次只使用一个样本更新模型参数，减少了每次计算的开销，适合于大规模数据集；另外SGD能够快速逃离局部极小值，适合于高维特征空间。

BGD适用于小规模数据集和低维特征空间，因为BGD需要使用全部数据计算梯度，计算开销较大，不适合于大规模数据集；另外BGD能够较快地收敛到全局最优解，适合于低维特征空间。

性能差异:

SGD的计算开销较小，每次更新只需要考虑一个样本，因此运行速度较快；但是由于更新过程的随机性，SGD的收敛过程不稳定，可能会出现震荡甚至跳出全局最优解的情况。

BGD需要使用全部数据计算梯度，计算开销较大，因此运行速度较慢；但是由于每次更新都考虑了全部样本，BGD的收敛过程相对稳定，能够较快地收敛到全局最优解。

因此，在选择优化算法时需要根据实际情况选择合适的算法，如果数据集较大，可以考虑使用SGD加速模型训练；如果数据集较小或者特征空间较低维，可以考虑使用BGD获得更好的收敛性能。此外，还可以使用一些综合了SGD和BGD优点的算法，例如小批量梯度下降（Mini-Batch Gradient Descent），它每次更新使用一小部分样本，既能够减小计算开销，又能够获得较好的收敛性能。

Q5 对于线性回归的最小二乘解， 什么情况会引起奇异解？ 具体表现有哪些？

在线性回归中，最小二乘法是一种常用的方法，它可以求得模型的闭式解（closed-form solution），也就是最小化残差平方和的解析解。然而，当数据存在一些特殊的情况时，最小二乘法可能会出现奇异解（singular solution）。以下是可能引起奇异解的情况及其具体表现：

数据不足： 当训练数据样本量少于特征数时，矩阵X的秩小于特征数，此时矩阵X的逆不存在，最小二乘法无法求解闭式解，可能出现奇异解。表现为残差平方和为0或者权重系数为NaN或Inf。

特征相关性：当训练数据中存在高度相关的特征时，矩阵X的秩可能小于特征数，此时矩阵X的逆不存在，最小二乘法无法求解闭式解，可能出现奇异解。表现为残差平方和为0或者权重系数为NaN或Inf。

数据噪声：当训练数据存在噪声时，矩阵X的秩可能会发生突变，从而导致矩阵X的逆不存在，最小二乘法无法求解闭式解，可能出现奇异解。表现为残差平方和为0或者权重系数为NaN或Inf。

当出现奇异解时，我们需要检查数据是否存在问题，例如检查数据的特征相关性、噪声等，或者使用正则化技术（如岭回归、Lasso回归）来缓解奇异解的问题。此外，也可以使用梯度下降等迭代优化算法来求解线性回归的解，这些方法通常不会出现奇异解的问题。

Q6 请简述线性回归最小二乘给出的封闭解和梯度下降给出的解之间的关系

在线性回归中，最小二乘法和梯度下降法都可以用来求解模型的参数。最小二乘法是一种解析解，可以直接得到模型参数的封闭解，而梯度下降法是一种迭代优化算法，通过不断更新模型参数，使得目标函数（如均方误差）逐步趋近于最小值。

封闭解可以直接求出模型的参数，不需要迭代求解，因此速度比较快，但对于数据集较大、特征较多的情况，计算矩阵的逆可能比较耗时和占用内存。而梯度下降法是一种迭代算法，需要不断迭代更新模型参数，直到达到一定的收敛条件，通常需要较多的迭代次数和计算量。

两者求解的结果是一样的，即最小化残差平方和，但是最小二乘法得到的封闭解可能会出现奇异解（如矩阵不可逆的情况），而梯度下降法通常可以避免这种情况。此外，梯度下降法具有一定的鲁棒性，可以处理带有噪声、异常值等不良数据的情况，而最小二乘法则对数据质量要求较高。

总的来说，最小二乘法和梯度下降法各有优劣，选择哪种方法取决于实际应用的场景和需求。

Q7 多项式回归和线性回归相比有什么优势？ 局限在哪里？

多项式回归和线性回归都是用于拟合数据的回归模型，二者的主要区别在于多项式回归可以处理更加复杂的非线性关系。

具体来说，多项式回归将自变量的高次项加入回归模型中，使得模型可以更好地适应非线性关系。这样，我们就可以用多项式函数来拟合数据，从而获得更加精确的预测结果。

多项式回归的优势在于可以适用于更加复杂的数据分布，可以更好地拟合非线性数据。此外，它的模型形式简单，易于理解和解释。另外，多项式回归还可以通过调整多项式的次数来控制模型的复杂度，从而平衡模型的偏差和方差。

但是，多项式回归也存在一些局限性。一方面，增加多项式次数会使模型变得复杂，可能会出现过拟合的情况；另一方面，高次多项式函数在较远离训练数据点的区域可能出现较大的波动，这意味着模型的泛化能力会受到限制。此外，多项式回归模型在处理数据时可能会出现数值不稳定的问题。

总的来说，多项式回归相对于线性回归具有更强的拟合能力，但在使用时需要注意过拟合的问题，还需要根据实际情况来选择多项式的次数。

Q8 除了归一化和z-score标准化， 机器学习中对数据的预处理还有哪些？

除了归一化和z-score标准化外，机器学习中还有一些其他的预处理方法，包括：

对数变换：将数据取对数可以使得数据更加符合正态分布，从而提高模型的拟合效果。

特征选择：选择最具有代表性的特征可以减少噪声和冗余信息的影响，从而提高模型的精度和泛化能力。

特征缩放：对不同尺度的特征进行缩放可以消除特征之间的影响，从而使得模型更加准确。

特征编码：将类别型特征转换为数值型特征，常见的编码方式包括独热编码和二进制编码。

数据增强：通过对原始数据进行变换和扩充，可以增加数据集的多样性，从而提高模型的泛化能力。

数据清洗：清除异常值、缺失值和重复值等错误数据，可以减少噪声的干扰，提高模型的鲁棒性和准确性。

不同的预处理方法在不同的场景下有不同的作用，需要根据具体情况进行选择和应用。