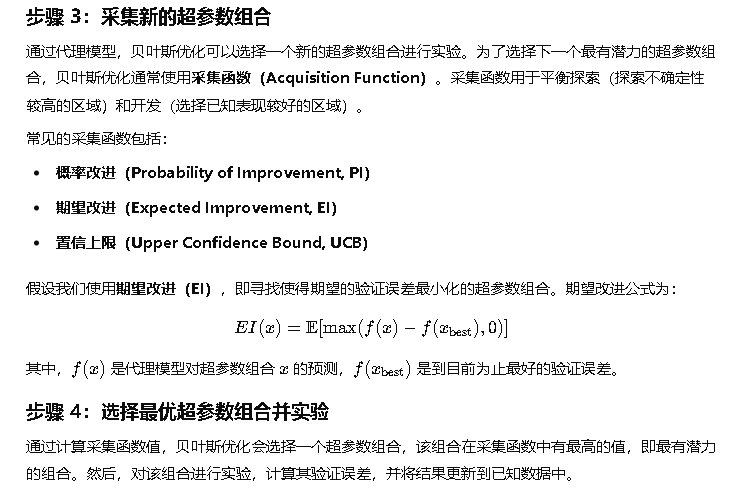
以下列举常见的无监督学习技术及其在异常检测中的应用：

#### ****降维与密度分析****

**方法原理**  
使用降维方法（如PCA、t-SNE）将高维数据映射到低维空间，异常点往往分布在低维空间中的边界区域。结合密度分析（如 LOF 或 DBSCAN），可以进一步检测低密度区域的点作为异常。

无监督学习在异常检测中扮演关键角色，其主要优势包括：

* 适用于数据标签稀缺的场景。
* 能够灵活适配各种数据类型（如图像、时间序列、文本等）。
* 提供多样化的方法（如聚类、降维、密度分析等）来发现异常模式。



**基于自编码器(1 from IEEE)的异常检测**

**[引用IEEE1目的:[我们使用它们来捕捉搜索空间的重要景观信息，然后在一个信息量很低的维度空间中产生后代。],**

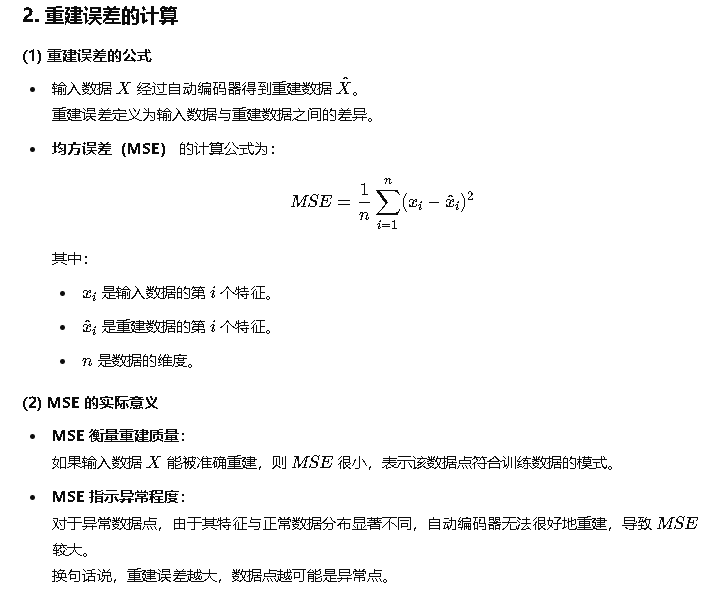
**引用IEEE2的目的:[(1)采用上述降维技术的目的是在降维空间中构建替代模型，这可以减少训练原始模型的时间复杂度并提高其准确性。[使用低维度特征的蛋白质和化合物表示，可以使用相对较少的参数，在后续进一步提取特征及相互关系。]**

**(2)在处理高维多目标优化问题（HEPs）时，传统的演化算法（EAs）搜索高维空间是困难的。自然地，我们采用自编码器作为降维工具，将原始空间压缩到一个缩减的空间，这对进化是有益的。**

**(3)自编码器的作用是基于训练数据提取有用信息，使得原始的高维搜索空间可以被转换到一个大幅度缩减的空间中，同时保留了有用的提取信息。此外，在缩减空间中的一个小搜索步骤，在原始空间中可能是一个巨大的步骤，这可能有助于找到高质量的高维多目标优化问题（HEPs）的解决方案。[]**

**(4)随着种群不断接近更好的区域，训练良好的自编码器可以提取有希望的进化方向的隐含信息，在接下来的优化中扮演着重要的角色。[将具有高维度特征向量的药物分子及蛋白质的编码，通过自编码器的编码(encode)，可以获取蕴涵了关键特征信息的、使用低维度特征向量表示的药物分子和蛋白质表示(representaion)]]**

**[进化计算解决的问题经常是数据获取稀少且极其昂贵的工程问题，因此需要先随机生成足够的数据来帮助自编码器学习[高维解到低维解的映射,这是为了方便在在低维景观中进行子代进化->以进一步方便快速的找到较优质的解],  
之后，分为两个子种群(进行种群合作以增强多样性),, 其中1个子种群借助代理模型进行更新,另一个子种群通过自编码器嵌入优化进行优化]]**



### ****3. 异常检测的逻辑****

#### ****(1) 训练阶段****

* 使用自动编码器对正常数据进行训练，使模型学习到正常数据的分布模式。  
  在训练结束后，自动编码器能够重建正常数据，而对于异常数据，其重建效果会较差。

#### ****(2) 重建误差的评估****

* 在检测阶段，计算每个数据点的重建误差（MSE）。
* 设定一个 **阈值**（如重建误差的95%分位数），将重建误差超过阈值的数据点标记为异常。

#### ****(3) 异常检测的关键假设****

* **正常数据分布：** 自动编码器能够很好地重建正常数据。
* **异常数据分布：** 异常数据远离正常数据分布，导致显著的重建误差。

#### 自编码器的用途： ****(2) 特征提取****

* 编码器的输出 Z (低维表示)通常被视为输入数据的特征表示（latent representation）。
* 这些特征可以用于后续任务（如分类或聚类）。

#### ****(4) 图像去噪****

* 将带噪图像作为输入，重建干净图像。

自编码器的常见变体

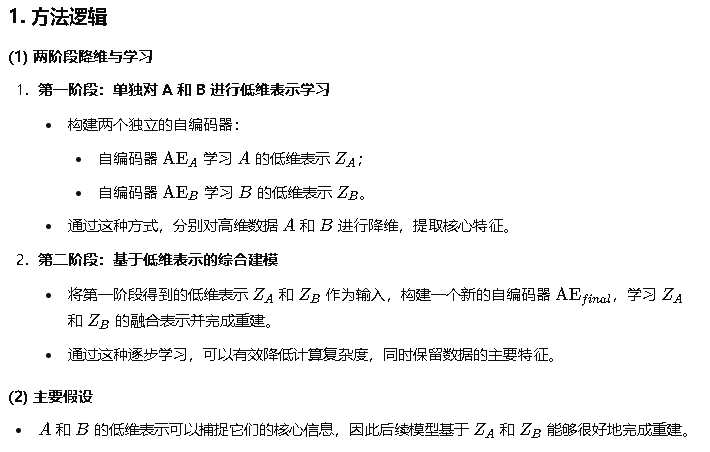
(5) 卷积自编码器（Convolutional Autoencoder, CAE）

将卷积层引入编码器和解码器，适用于图像数据的重建和特征提取。

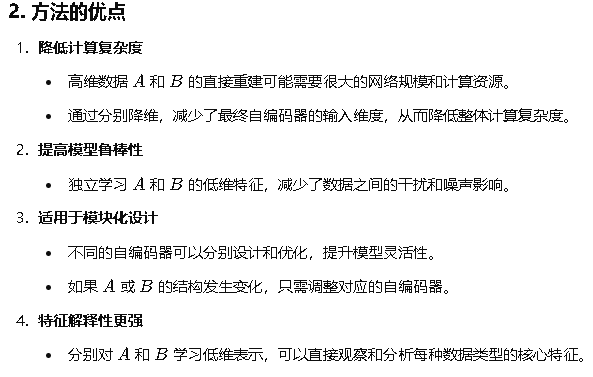
(6) 对抗自编码器（Adversarial Autoencoder, AAE）

将生成对抗网络（GAN）的思想融入自编码器中，用于生成更逼真的数据。

上述自编码器中使用的信息A及B的维度是巨大的，可否在使用原有的自编码器之前，先使用另外的2个自编码器，分别学习A及B的低维度表示。后续再使用原有的自编码器时，就只需要使用A,B的低维度表示即可，而不用使用其高维的表示。



关键点：



对于B的自编码器的学习。由于B数据有600组。但是B的每组数据的序列长度不一定相同。例如，第1组数据的形状为(100,1500),第2组数据的形状为(136,1500),第3组数据的形状为(400,1500)，等等。请提供给我新的对于B的自编码器，要求是，该自编码器在训练时，在每个epoch内，将每组数据都分为若干长度不同的子序列，并使用注意力机制学习这若干子序列之间的关系。在后续使用自编码器时，输入形状为(n,1500)的数据时，可以获得输出为(n,30)的低维度表示。

这种需求中，B 数据具有变长序列特性，因此需要设计一种 基于注意力机制的自编码器，以便适应变长输入，并提取子序列间的全局关系。

