

PB22111645朱恩松

1.实验流程

按实验文档顺序进行即可。

2.调参过程

未调参时聚类效果不佳，考虑到降维后维数越低信息越少，故提高降维后的维数，发现效果良好。

3.结果

见results文件夹。

4.回答问题

1

- 1.训练速度：PCA快，AutoEncoder次之，tSNE最慢。
- 2.降维效率：三者差距不大。
- 3.灵活性：PCA仅支持线性降维，灵活性差；tSNE专注于非线性降维，但仅适用于2D/3D；AutoEncoder灵活性最强，可处理各种类型的数据，能够降到任意维度，并支持对数据分布的非线性建模。
- 4.对数据分布的保持程度：PCA只能捕捉数据的线性结构，对数据的全局结构保持较好；tSNE能很好地保持数据的局部结构，但对全局结构保持较差；AutoEncoder能同时保留数据的局部和全局结构，适用于复杂非线性分布的降维。
- 5.可视化效果：PCA效果一般，生成的低维数据易于解释，但只能处理线性特征，可能忽略非线性关系，适合简单分布的数据；tSNE效果优秀，能很好地展示高维数据的局部结构，使聚类 and 分类结果更加直观；AutoEncoder可视化效果依赖解码器的设计和训练结果。

2

- 1.生成效率：GMM生成样本直接从高斯分布采样，效率高；DDPM通过多步逆扩散过程生成样本，生成一个样本可能需要数百到上千次迭代，效率较低。
- 2.生成质量：GMM假设数据服从高斯分布，适合低复杂度数据，对于复杂分布处理能力有限；DDPM能生成高质量样本。
- 3.灵活性：GMM仅适用于结构化、低维数据，难以建模复杂分布。DDPM适用于各种类型的数据，对复杂数据分布有强大的建模能力。
- 4.是否可控：GMM生成指定类别样本非常简单，只需从对应类别的高斯分布中采样即可，具有明确的类别控制能力；DDPM通过添加条件输入可以实现类别控制，但需要模型在训练阶段已加入条件生成机制，增加了实现和训练的复杂度。

5.反馈

书面作业一天，实验一天。