**数据科学基础第一次作业**

1. **维度灾难**

分别从维度为2，100，1000，10000的标准正态分布采100个数据点，计算所有数据点之间的距离，并绘制这些距离值的直方图，实验结果参考图 1。

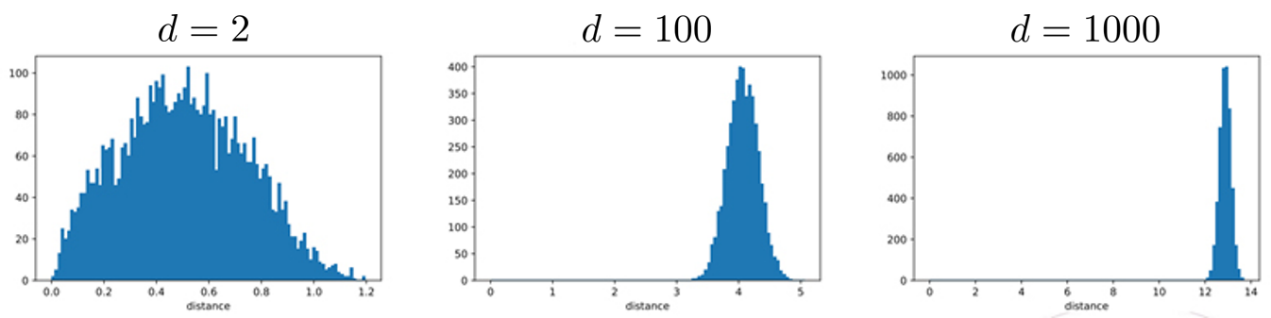


图 1

1. **数据降维**

**数据集介绍**：数据集包含了三个被试在积极，消极以及中性情绪状态下脑电信号的微分熵特征，每个数据样本都包含62通道的信号，每个通道又划分为5个频段，因此每个样本点的特征有310维（数据集下载链接<https://cloud.bcmi.sjtu.edu.cn/sharing/fmcQ3cewh>）。数据集的每个文件夹下包含了单个被试的数据，总共三个被试（文件夹名分别为1，2，3），训练数据和测试数据已划分好（0：消极，1：中性，2：积极）。

**实验要求**：

（1）分别用PCA, LDA, TSNE三种降维方法对单个被试的数据降维（降至2维），并可视化降维后的数据（不必给出每一个被试的结果，给出任意一个被试降维后的数据可视化结果即可）。

（2）分别用PCA, LDA, TSNE三种降维方法对所有被试的数据（将所有被试的数据结合在一起）降维（降至2维），并可视化降维后的数据。

（3）使用SVM和KNN实现**被试依赖**的情绪分类（使用原始数据，即未降维的数据）。（被试依赖：在训练模型时，只使用被试自己的数据做训练集和测试集，也就是说在本次实验中，对三个被试要分别输出分别训练三个SVM和KNN）

作业要求：需要提交实验报告和代码（代码实现允许调第三方库），实验报告至少需要包括基本的实验结果，请将所有需要提交的文件打包成压缩包，命名格式为“姓名\_学号\_第一次作业”，并于10月25日23:59前提交至canvas平台。