实验 5：隐私保护实验指南

一、 实验目的

学习差分隐私的原理，使用差分隐私进行隐私保护的实践。

二、 实验步骤

1. Python 环境准备

首先需要配置一个 Python 环境，并且需要以下包：numpy, pandas, pytorch 或 tensorflow。

2. 实现一个网络对 MNIST 数据集分类

提供的样例文件里已经有了一个基于逻辑回归的分类器，大概可以达到 90%的分类准确率。请实现一个更复杂的网络，使得验证集上分类准确率到 95%以上。

3. 对模型输出结果进行差分隐私扰动

通过对**模型输出结果**添加噪声，我们可以使得模型输出结果具有ϵ的差分隐私，从而保护模型的信息。同时，考察添加噪声后的预测值的验证集分类准确率。

4. 分析差分隐私机制对准确率的影响，尝试给出数学推导

阅读差分隐私资料，理解差分隐私机制，**说明差分隐私的ϵ和最终结果的准确率的关系**（绘制曲线图）。

三、 提交内容

实验报告：包含上述实验内容（实验 4 可选）。理论推导可 latex 编辑、word 或者手写拍照。

实验代码：模型代码和差分隐私机制的代码

参考文献：

差分隐私学习资料：

The Algorithmic Foundations of Differential Privacy