癌症乃是人类健康的一大杀手，对于癌症的分析和预测也成为现代医学领域中的一个重要问题。我们希望使用一批患者的检查样本数据（附件中的 breast_cancer.csv 文件），通过建立一个逻辑回归模型来学习这些样本数据，来得到一个较好的乳腺癌预测模型。

数据一共有 569 组 31 列，其中前面 30 列是对于患者乳腺部位的检查情况（包括 mean radius、mean perimeter 等特征），最后一列 type 表示是否患有乳腺癌（0 表示不患乳腺癌，1 表示患乳腺癌），共有 212 个正样本（type=1）和 357 个负样本（type=0）。

1. 读取数据（可使用 pandas 的 read_csv 函数读取 csv 文件），划分训练样本和测试样本，搭建逻辑回归模型，并计算在测试集上预测的准确率。
2. 计算测试集上预测结果的混淆矩阵，填充如下表格。

	预测不患癌症	预测患癌症
实际不患癌症	a	b
实际患癌症	c	d

3. 输出逻辑回归模型的参数 k0-k30，对每一个**测试样本**计算对应的 y 和 f（y）值，画出 y 与 f（y）的散点图，其中正样本以红色表示，负样本以蓝色表示。

$$y = k_0 + k_1 x_1 + k_2 x_2 + \cdots + k_{30} x_{30}$$

$$f(y) = \frac{1}{1 + e^{-y}}$$

编写程序完成上述要求，并写一份 500 字左右的实验报告（pdf 格式，包括程序实现说明，实验结果及所画图像，实验结果分析），2021 年 4 月 30 日前将源代码与实验报告提交到 chenty@stu.pku.edu.cn