- 1. 读取 breast-cancer-wisconsin.csv 文件
- 2. 输出前五行查看数据内容
- 3. 将数据中的缺失值删除(提示:将?替换为 nan, 去掉 nan 所在的行)
- 4. 划分数据集为训练集和测试集(其中第 1 个列为 id 号,第 2-10 列为特征,11 列为标签(2 为良性、4 为恶性))
- 5. 对数据进行标准化处理
- 6. 使用逻辑回归进行预测
- 7. 输出训练集和测试集的错误率