模型优缺点与改进

二元logistic回归模型（第三问模型）

优点：模型简单，速度快，适合二分类问题；

简单易于理解，直接看到各个特征的权重；

能容易地更新模型吸收新的数据；

缺点：学习能力较弱，对数据和场景的适应能力有局限性；

容易忽视变量之间的相关性，在强相关的因素下，一些弱相关的因素会被淹没；

不患病和患病居民数量级差距太大，造成样本特征被淹没。

模型改进：

1. 用主成分分析进行数据降维，将几个大类用主成分分析归为部分抽象变量看待。
2. 给慢性病按照病情严重程度分3档或者5档，增加模型预测能力。
3. 使用SMOTE（合成少数类过采样技术）分析少数样本并人工合成新样本添加入数据集，以缓解类别不平衡影响的模型输出