Y = b0 + b1x1 + b2x2 + … + bnxn

Logistic regression:

1. 预处理数据集
   1. 训练集：历史上所有比赛+世界杯2014年前比赛 labeled minimize loss function
   2. 验证集：2018年世界杯比赛.
2. 数据集的处理：
   1. 提取出世界杯参赛球队的数据
   2. 例子：预测阿根廷和沙特阿拉伯
      1. 从数据集中找到阿根廷和阿拉伯这两支球队的在1930之后2014年之前的所有比赛（信息特征提取）去掉friendly
      2. 从数据集中找阿根廷和阿拉伯2014-2022年的所有比赛 （训练集）#变量名是什么，head（） 最晚周三 pandas ZHUOQUN, Panda dataframe
      3. 计算阿根廷（阿拉伯）最近15场次比赛的胜率 x1 float [0, 1]
      4. 计算阿根廷和阿拉伯最近5场比赛胜负情况（用一个数字表示）x2 int [-5, 5] Mai
      5. ~~提取出阿根廷（阿拉伯）最近15场比赛的季节 x3 []~~
      6. 提取出阿根廷（阿拉伯）最近15场比赛净胜球 x4 int [-Inf, Inf] 最晚周四 Yu
      7. 建立模型z = b0 + b1x + … y = sigmoid(z)
      8. 会有五个参数 b0, b1, b2, b3, b4
      9. 我们用ii的训练集来训练b0 ~ b4
      10. 用训练好的参数，计算2018世界杯这场比赛的胜率，用结果验证
      11. 用训练好的参数，算2022世界杯结果 最晚周五 Dongxiao
3. Google colab jupyter notebook