

建模思想

本题中提供了球赛对局中的多个参量，且为非线性关系，并不能直接通过线性回归等方法得出结论，故考虑使用卷积神经网络（CNN）进行多分类，对参量与目标量进行深度学习训练，得出一定的结论。

对CNN及其使用的简述

卷积神经网络：CNN

卷积神经网络（Convolutional Neural Networks，简称 CNN）是一种深度学习模型，主要用于处理图像、语音或音频等多维数据。卷积神经网络的特点是利用卷积层、池化层和全连接层等组件，从输入数据中提取特征，通过梯度反传不断深化学习，并进行分类或回归等任务

本题使用

- 输入输出数据的结构
  - 输入：7284\*1\*12，12为特征数，7284为数据量，1为通道数；
  - 输出：7284\*27，27为分类数，7284为数据量
- 网络结构：卷积核为1个
  - 卷积层1，采用ReLU激活函数，以下卷积层与线性连接层使用与此相同的激活函数
  - 最大池化层1:提取主要特征，即突出特征
  - 卷积层2
  - 最大池化层2:提取主要特征
  - 卷积层3
  - 线性连接层1
  - 线性连接层2
  - 线性连接层3
  - 归一化层（输出层）
- 特殊处理：在输出层最后一层加入Softmax归一化层，可以使得数据分布在0，1之间，便于分类
- 损失函数：交叉熵损失函数
- 优化器：采用性能较好的Adam优化器
- 评估检验：采用Accuracy,F1\_score,Recall,Precision四项检验指标

建模过程

C题第一问

- 指标选取：采用状态点的建模策略，状态点内容包括“已进行的局数，是否双误，.....”等其他指标的集合。对状态点进行分类学习。
- 目标选取：选用连续赢球数作为优势指标，以球员1连续赢球数减球员2连续赢球数为相对优势指标并量化。
- 建模：对数据进行预处理后进行CNN训练，采取随机分割数据集（3:1）的方法，保证随机性，突出主要特征，减少过拟合风险
- 绘图：对test数据集进行模型分类估计，绘制"x\_state-原数据，预测值"的散点图，其中x\_state为从0到7283的整型序列

C题第二问

去掉球员的已赢盘数，局数，分数之差，对神经网络进行重新训练，召回率从94.5%左右降至93%，由于召回率的特性，可认为连续胜利对球员优势有影响（存疑🤔）