1. **确立合适的任务目标**

由于每只股票每天的行数是一直在变的，在时间上的分布也往往是不均匀的，因此直接拿前x行对后y行做回归是不合理的。合理的方式是根据前x天的数据，对后y天的特征分布做回归。

那么，不妨假设目标任务是根据前5个交易日的特征数据来对接下来的1个交易日的特征数据的分布做回归，回归目标有下来的1个交易日的均值、中位数、行数、标准差、最大值和最小值。

1. **数据处理**

首先，将数据按照每只股票分组。接下来，由于每只股票每天的行数为1000~5000行，因此将每天的时间均分成1000个时间格，根据特征的时间戳，将这些行的特征分别放进对应的时间格中，如果有多个特征的时间戳都位于同一个时间格，那么将这些特征的值取平均（如果是价格类的特征，则取平均；如果是成交量类的特征，则取求和）后放入对应的时间格中。如果有些时间格没有特征，则赋值为0。

这样，每只股票每天的数据就可以变成的矩阵。

接着，将每只股票每天的原始数据的统计量计算出来，统计量为均值、中位数、标准差、最大值、最小值和行数，这些统计量作为回归任务的label，是一个的矩阵。

那么训练数据和label就都有了，训练数据为每只股票连续5天的特征矩阵，形状为；对应的label为每只股票接下来1天的统计量矩阵，形状为。

1. **模型训练**

模型：使用1-D CNN作为模型，模型有1-D卷积、RELU和FC-layer（个concatenate在一起）。

损失函数：MSE

Optimizer：Adam

流程图：

