

数据集说明：

给定训练集 train.csv，要求根据前 9 个小时的空气监测情况预测第 10 个小时的 PM2.5 含量。

训练集介绍：

- (1)、CSV 文件，包含台湾丰原地区 240 天的气象观测资料(取每个月前 20 天的数据做训练集，12 月 X20 天=240 天，每月后 10 天数据用于测试，对学生不可见);
- (2)、每天的监测时间点为 0 时，1 时.....到 23 时，共 24 个时间节点;
- (3)、每天的检测指标包括 CO、NO、PM2.5、PM10 等气体浓度，是否降雨、刮风等气象信息，共计 18 项

用 excel 打开，繁体字会出现乱码：

用 notepad++可以正常打开：

作业说明：

根据作业要求可知，需要用到连续 9 个时间点的气象观测数据，来预测第 10 个时间点的 PM2.5 含量。针对每一天来说，其包含的信息维度为(18,24)(18 项指标, 24 个时间节点)。可以将 0 到 8 时的数据截取出来，形成一个维度为(18,9)的数据帧，作为训练数据，将 9 时的 PM2.5 含量取出来，作为该训练数据对应的 label；同理可取 1 到 9 时的数据作为训练用的数据帧，10 时的 PM2.5 含量作为 label.....以此分割，可将每天的信息分割为 15 个 shape 为(18,9)的数据帧和与之对应的 15 个 label。

训练集中共包含 240 天的数据，因此共可获得 $240 \times 15 = 3600$ 个数据帧和与之对应的 3600 个 label。

将前 2400 个数据帧作为训练集，后 1200 个数据帧作为验证集。

输入：

训练集 train.csv

输出：

- 1.训练集每轮 epoch 的 loss;
- 2.验证集的 loss。

样例：

```
after 600 epochs, the loss on train data is: 44.88913061600439
after 800 epochs, the loss on train data is: 44.26903588227097
after 1000 epochs, the loss on train data is: 43.950109190566856
after 1200 epochs, the loss on train data is: 43.78092633274224
after 1400 epochs, the loss on train data is: 43.68982565130423
after 1600 epochs, the loss on train data is: 43.640314303297686
after 1800 epochs, the loss on train data is: 43.61322589236443
The loss on val data is: 40.35422383809947
```