# 实验记录

## 1时间：20190611，00:00

## 2实验目的：

测试网络是否可以运行，测试新整理得到的数据是否可以进入网络，针对网络进行调整。

## 3实验代码存放文件夹：

20190611文件夹中。

## 4实验运行时间：

1天及以上。

## 5实验使用数据信息：

药物作用位点预测，可能存在的药物作用位点标志为1，正样本。其余其他位点均为负样本。

根据6月6日之后整理得到的数据，正样本选取所有正样本点，去同（去掉所有一样的）

负样本来自0.3去冗余后的蛋白质中，去掉处于正样本中的。

## 6.1，数据集正负样本设置，比例设置，

正负样本比大概1：12，按照8:1:1划分训练集、验证集、独立测试集。随机打乱后划分。

训练中随机选取正样本中80%的数量进入训练。取相同数量的负样本。

## 6.2，网络设置：

使用何飞老师的网络

增大filter\_size, 扩大5倍，扩大10倍

增大步长，步长为10，步长为20

试验原本模型对比

使用one hot 编码

构建全连接网络，

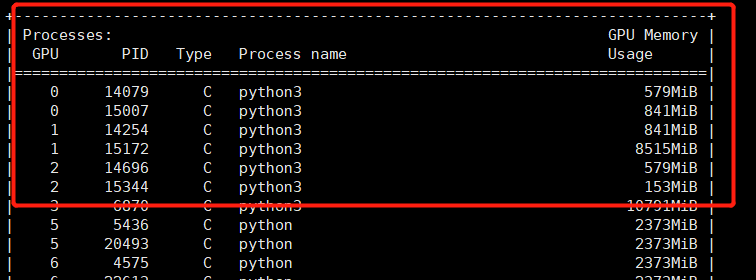
构建2维卷积网络。

查找已有网络

## 8实验运行记录

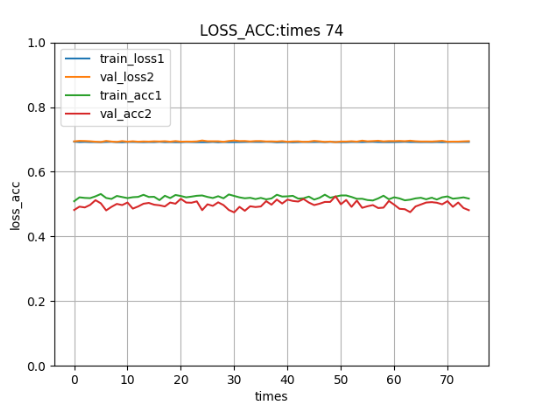
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 开始时间 | 输出文件 | 结束时间 | GPU编号（数量） | 运行id | 文件夹 |  | 步长 | Kernal size | 描述 |
| 1 | 23：00 | 如图 |  | 0 | 如图 | 0 | step\_10\_size\_5\_gpu0\_result\_0 | 10 | 扩大5倍 |  |
| 2 | 23：00 | 如图 |  | 0 | 如图 | 1 | step\_10\_size\_10\_gpu1\_result\_1 | 10 | 扩大10倍 |  |
| 3 | 23：00 | 如图 |  | 0 | 如图 | 2 | step\_20\_size\_5\_gpu2\_result\_2 | 20 | 扩大5倍 |  |
| 4 | 23：00 | 如图 |  | 0 | 如图 | 3 | step\_20\_size\_1\_gpu3\_result\_3 | 20 | 扩大10倍 |  |
| 5 | 23：00 | 如图 |  | 0 | 如图 | 4 |  | 原始 | 原始 | 矩阵转置，row-col=1-2 |
| 6 | 23：  00 |  |  |  |  | 5 |  | 原始 | 原始 | 矩阵不转置，row-col=2-1（原本） |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

以后根据预期主要变量由小到大设置数值，GPU编号，名字，文件夹都要由小到大设置。

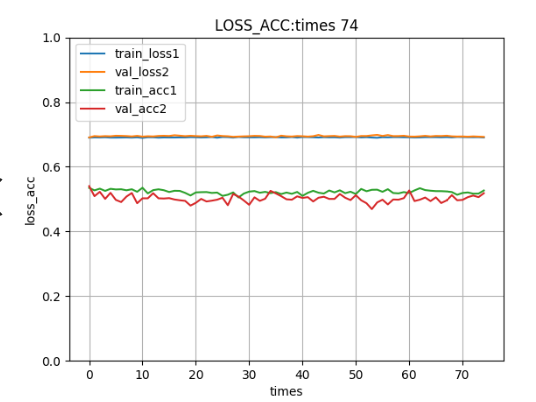


## 9实验结果存放位置

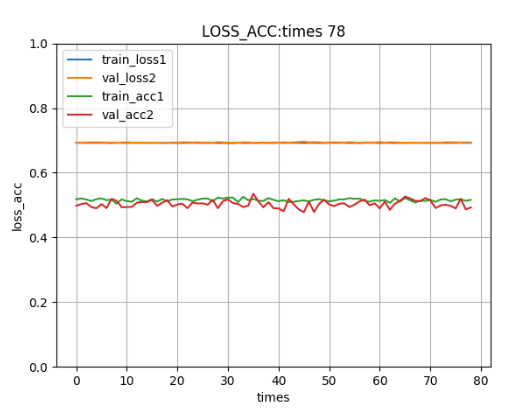
## 10结果分析



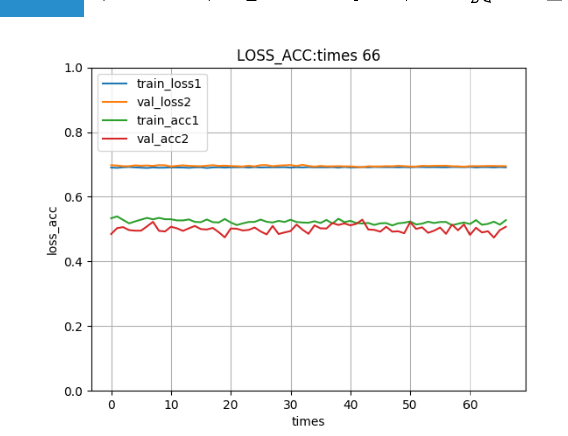
实验1，步长扩大10倍，卷积核扩大5倍，acc有变化。上下波动。



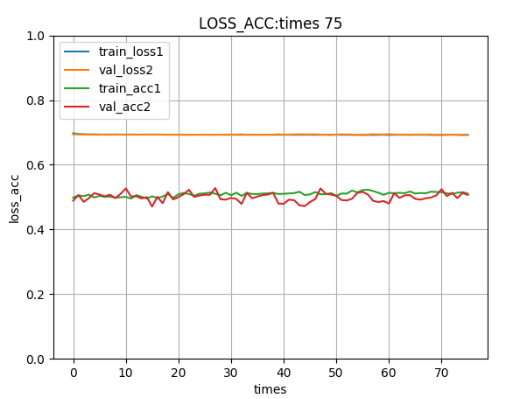
实验2，步长和卷积核大小分别增大10倍，结果有起伏。但观察卷积核的改变没有显著影响。



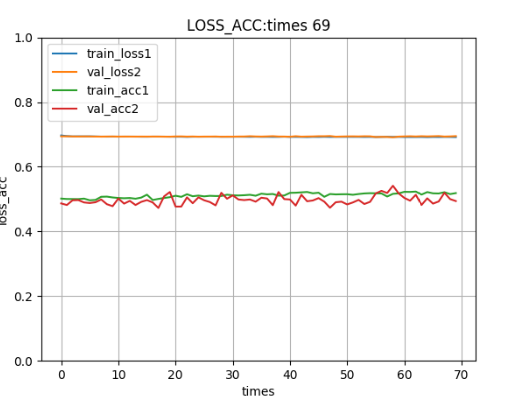
实验3，步长扩大20倍，卷积核扩大5倍，除了提升了一点运行速度没啥变化。



实验4，步长扩大20倍。卷积核扩大10倍，运行减慢了8个。



步长和卷积核大小没有变化的情况是这样的。



经过重新输入shape的结果是这样的。根据今日的学习，不管什么shape,最终结果都会是长度1维序列。

只要一维序列没有变化，则机器不能发现我们的输入有问题。

## 

## 11总结

1网络随意尝试效率太低。下次调整网络，计划找一个效果好的。或者根据先验知识调整。

2数据是关键，今天发现数据问题比较严重。也很心疼自己的网络：先告诉它这个是正样本，一会儿又告诉它这个这个是负样本。。。假阴性样本多？