深度学习对比实验

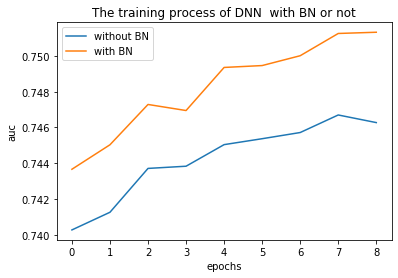
以下实验均进行三次重复实验取均值作为最终结果。

DNN模型规模：

Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

对比方式：是否BN

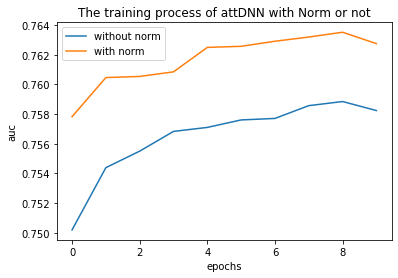


attDNN模型规模：

Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

对比方式：是否softmax归一化



attDNN模型规模：

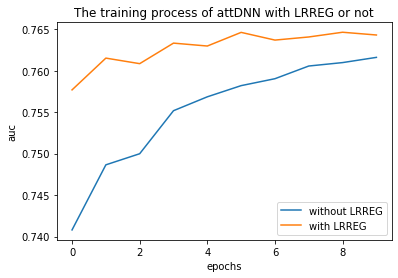
Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

归一化方式：softmax归一化

其它层：添加LR层

对比方式：是否对LR进行正则化



attDNN模型规模：

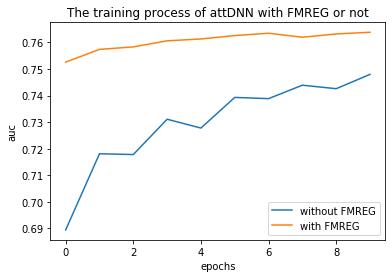
Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

归一化方式：softmax归一化

其它层：添加FM层

对比方式：是否对FM进行正则化



attDNN模型规模：

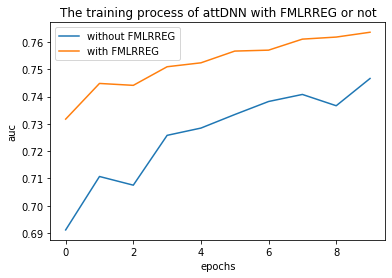
Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

归一化方式：softmax归一化

其它层：添加FM层以及LR层

对比方式：是否对FM、LR进行正则化



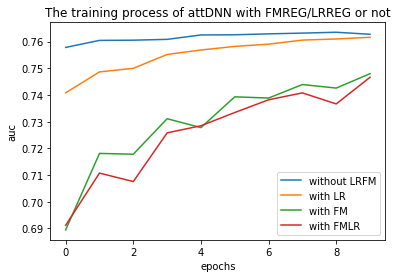
attDNN模型规模：

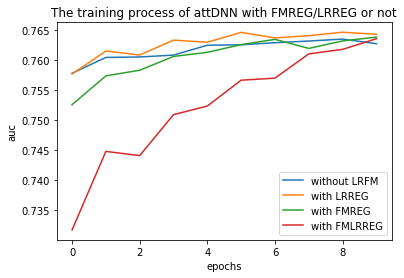
Embedding维度：city+aid -> 16、其他 -> 8

隐层规模：512 –> 256 -> 128，relu激活函数

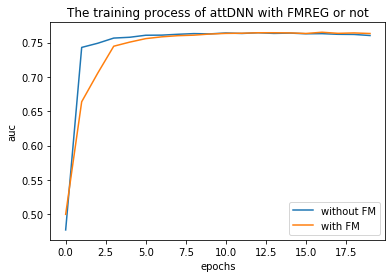
归一化方式：softmax归一化

对比方式：是否添加FM、LR以及都添加





为了区分FM的作用，进行是否添加FM模块的实验。



通过上面的重复试验，发现以下结论：

1. BN能够促使深层前馈网络模型收敛更快，效果更好；
2. 加入softmax归一化后的Attention才能起到提升模型效果的作用；
3. FM、LR模型在加入参数正则化后更能对模型有提升作用；
4. 正则化后的FM加入前后对整体模型影响不大。

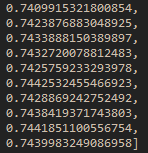
因此，初步较优模型应该是以下组合模型结构：

BN + softmax归一化Attention + 正则化LR

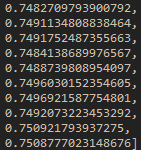
下面是单次实验的结果截图。

# 无label权重，普通DNN

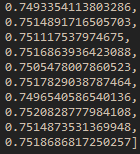
## 3epoch 32dims 128-64-32



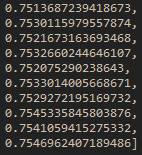
## 5epoch 32dims 128-64-32



## 7epoch 32dims 128-64-32



## 5epoch 32dims 512-256-128

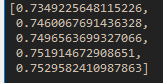


# 有label权重，普通DNN

## 5epoch 32dims 512-256-128



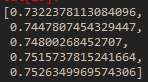
尝试性，由于各个特征下对应的种类不一致，为了减少Embedding维度过长带来的过拟合影响，调整维度为aid+city保持32，其他变为16



后面的实验调整维度为aid+city：16，其他变为8



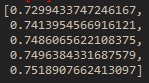
添加了gametype的偏置项



# 无label权重，AttDNN

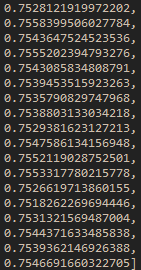
## 5epoch 16dims-8dims 512-256-128

1.未归一化，单纯Attention权重乘以weight





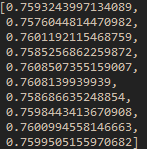
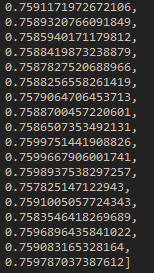
10epoch（验证是否可以一直上升）



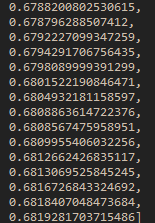
不会一直上升

2.Softmax归一化

一次实验和二次实验



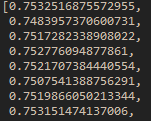
3.后续想要尝试设置不同的待训练的Attention向量来针对不同的序列特征进行Attention计算，已尝试直接将Attention融合后的特征向量拼接到dense输入层中，效果不理想，下图即效果，继续尝试将新的Attention权重作为乘子添加到原生Attention模块中，再继续归一化查看效果，发现添加前后效果相似，因此剔除该固定Attention权重。



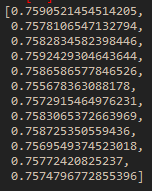
# 无label权重，归一化AttDNN，添加FM|LR

以归一化模型的0.76AUC作为baseline

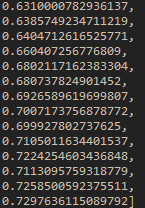
## 只添加FM层（只有二次特征交互）



## 只添加LR层



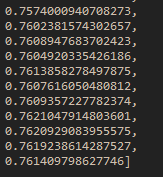
## 添加FM层&LR层



收敛慢+效果降低

## 添加正则化层和bn

添加正则化



添加bn（swish）

