Extensive Linear Connectivity and Application in Model Ensemble

1. Introduction

1.1 mode connectivity及两种连接方式

mode connectivity的相关研究试图在两个minima对应的参数θ_1, θ_2（minima对应的参数θ_1,θ_2被称为mode）之间寻找一个low loss的路径，在寻找路径时，最直接的想法是直接考虑连接θ_1和θ_2的线段，但是在大多数情况下，连接θ_1和θ_2的线段上的loss并不总是低的，会存在一个较高的loss，被称为barrier。所以，在寻找连接参数θ_1, θ_2的low loss的路径时，并不会直接考虑直线路径，当前研究主要考虑**曲线路径**[1] [2]或者**置换后的直线路径**[3]。

曲线路径是指，对于任意的mode θ_1,θ_2，计算曲线 $\phi_{\theta_1,\theta_2}(t), t\in[0,1]$，$\phi_{\theta_1,\theta_2}(0)=\theta_1, \phi_{\theta_1,\theta_2}(1)=\theta_2$使得$\forall t\in[0,1]$，参数$\phi_{\theta_1,\theta_2}(t)$对应的loss都很低。在[1]中，通过类似于物理中的Nudged Elastic Band过程计算出了这样的low loss曲线，而在[2]中，将曲线$\phi_{\theta_1,\theta_2}(t)$的类型限制为折线（一般为一折的折线）或者贝塞尔曲线，然后通过优化方法计算出控制折线和贝塞尔曲线的参数。

在考虑置换后的直线路径时，需要注意的是，神经网络满足一定的置换性。考虑M层全连接神经网络，神经网络每一层大小为$n_0,\dots,n_M$，其中第0层为输入层，第M层为输出层，设参数θ_i对应的每一层的参数为矩阵 $W_j^{(i)}\in R^{n_j\times n_{j-1}}, 1\le j\le M$（为了叙述简便，这里不考虑神经网络偏置$b_j$，实际上讨论类似），第$j$层经过激活函数过后的结果为$z_j^{(i)}=\sigma_j(W_j^{(i)}z_{j-1}^{(i)})$，$z_0^{(i)}$表示输入向量。对于任意的置换矩阵序列 $P=\{P_0=I,P_1,\dots,P_M=I\}$，对神经网络$\theta_i$进行置换变换得到的新的神经网络$P(\theta_i)$，$P(\theta_i)$每一层的参数定义为$W_{j,\text{new}}^{(i)}=P_jW_j^{(i)}P_{j-1}^T, 1\le j\le M$。注意到，新的神经网络$P(\theta_i)$表示的函数和神经网络$\theta_i$表示的函数是同一个函数，这个不变性被称为神经网络的permutation symmetry，而变换$P(\cdot)$被称为神经网络的一个对称变换。

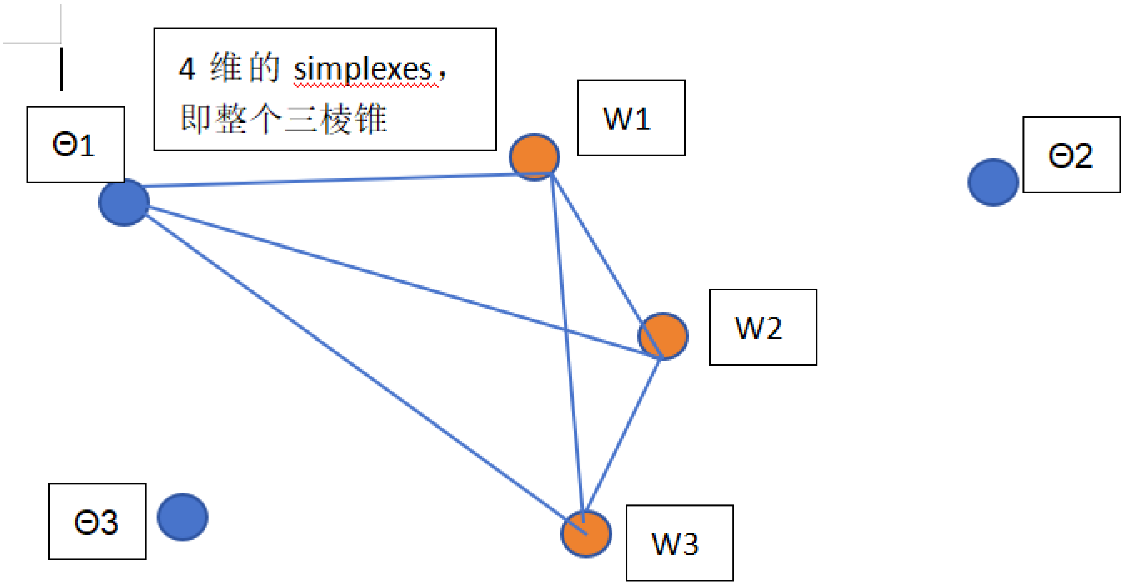
置换后的直线路径是指，对于任意的mode θ_1,θ_2，尽管θ_1和θ_2不存在low loss的直线路径，但是存在一个对称变换P，使得θ_1和$P(\theta_2)$之间存在low loss的直线路径。[3]给出了三种方法用于计算这样的变换P。

需要注意的是，这两种连接方式之间不存在包含关系！！！！！！！！

它们是两种完全独立的连接方式，这也是为什么说曲线连接和直线连接不同的原因（直线连接指经过置换后的直线路径）

1.2 曲线路径连接的扩展：simplexes connectivity

曲线路径的mode connectivity可以扩展到mode simplexes connectivity[4]，在[4]中，将mode之间曲线的连接扩展成了低维数的Simplicial complex的连接（在[4]中，实验验证了进行连接的Simplicial complex的维数一般低于10），simplexes connectivity的具体描述如下为：对于mode $\theta_1, \theta_2, \ldots, \theta_i$，存在网络的参数 $w_1, \ldots, w_k$(k<10)，使得$\forall 1 \leq j \leq i$,$w_1, \ldots, w_k, \theta_j$ 构成的单纯型（simplex）上的loss都很低，其图示为：



simplexes connectivity可以视为[2]中折线形式的曲线路径mode connectivity连接的扩展，在[2]中折线形式的曲线路径实际上是k=1,即一个网络参数w_1，然后使得单纯型w_1,θ_1和单纯型w_1,θ_2上的loss 很低。所以，这个想法实际上是对曲线路径的一个扩展，我期望将它扩展到**置换后的直线路径**上去。

1.3 置换后的直线路径连接的扩展：simplexes connectivity

置换后的直线路径来说，需要首先注意到两个性质

性质1：考虑任意对称变换P，以及mode θ_1,..., θ_n,参数$k_1, \ldots, k_n$,参数$\sum_{i=1}^n k_i \theta_i$和$\sum_{i=1}^n k_i P(\theta_i)$分别对应的神经网络表示同一个函数

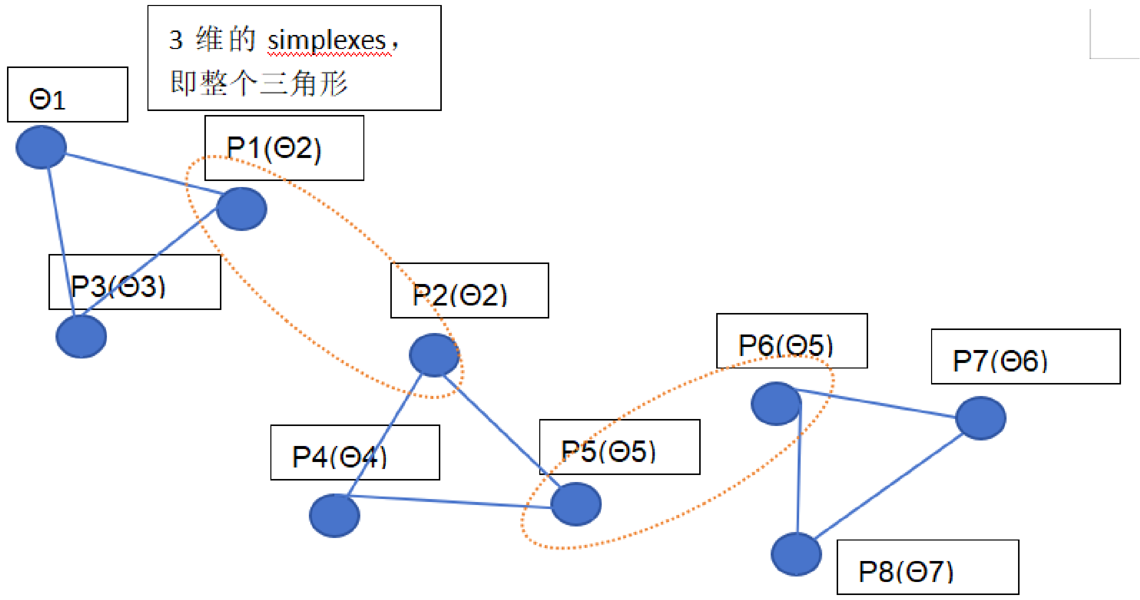
性质2：对于mode $\theta_1, \theta_2, \theta_3$，有[3]中的算法知，存在对称变换$P_1$,$P_2$，使得$\theta_1, P_1(\theta_2)$之间和$\theta_1, P_2(\theta_3)$之间分别存在low loss的直线路径，但是[3]中的算法并不保证$P_1(\theta_2), P_2(\theta_3)$之间存在low loss的直线路径，更不保证，$\theta_1, P_1(\theta_2), P_2(\theta_3)$构成的单形上任意一个参数点是low loss的。

在下面的叙述中，记$\bar{\theta_i}$为等价类
$\{P(\theta_i)|\text{对于任意对称变换P}\}$。

由于变换P的高维性，从我的直观上来看，对于一定数量的mode $\theta_1, \theta_2, ..., \theta_n$,可能存在置换$P_1, ..., P_{n-1}$使得$\theta_1, P_1(\theta_2), ..., P_{n-1}(\theta_n)$组成的单纯型上任意一个参数点是low loss的，但是和[4]中simplex的维数限制一样，这个n可能不会太大。

这里和曲线路径的simplexes connectivity有一个不同，对于mode $\theta_1, ..., \theta_n$，曲线路径的simplexes connectivity，不是使得$\theta_1, ..., \theta_n$中某些mode组成的simplex上每一点是low loss的，而是找到了一些其他的参数点$w_1, ..., w_k$使得$\theta_j, w_1, ..., w_k$组成的simplex上每一点是low loss的。而在置换后的直线路径的simplexes connectivity中，我们是找到置换$P_1, ..., P_{n-1}$使得$\theta_1, P_1(\theta_2), ..., P_{n-1}(\theta_n)$组成的单纯型上任意一个参数点是low loss的。

上面提到，在置换后的直线路径的simplexes connectivity中，组成low loss simplex的参数点的数量会比较小，所以我们需要将这些simplex粘合起来，形成一个更庞大的Simplicial complex，粘合的方式非常自然，即通过等价类即可，如下图所示：



虚线的圆圈表示这个圆圈中的节点属于同一个等价类，所以可以将它们视为同一个点，在这个意义上，这些simplex粘合成了一个Simplicial complex。这样，通过对称变换，定义了一个mode之间可能新存在的simplex connectivity的方式。这个simplex connectivity是置换后的直线路径的connectivity的扩展。

2. Possible Algorithm

3. Related Work

[1] Essentially No Barriers in Neural Network Energy Landscape

[2] Loss Surfaces, Mode Connectivity, and Fast Ensembling of DNNs

[3] Git Re-Basin: Merging Models modulo Permutation Symmetries

[4] Loss Surface Simplexes for Mode Connecting Volumes and Fast Ensembling