

Lora合并

- 关于Boosting, Boosting的本质是弱分类器来实现“强”分类效果, 但是Boosting是需要去训练分类器的, 而我们的lora是已经训练好的了
- 因此我们可以接着Boosting的概念进行包装, 虽然不太搭嘎, 但是我们可以说 $W = xW_1 + yW_2 + zW_3$ 这个东西里, 每个分项整体都是弱分类器, 我们要训练的是前面的权重参数
- 上面的实现, 我们可以手动用传统一层神经网络来实现
- 如果只是上面的做法, 可能本身厚度不够, 可以接着吹嘘并引入“辅助网络”的概念
 - 辅助网络这个名词应该不准确, 其对应的概念是在最初训练一个较大的神经网络的时候, 由于太大了, 导致计算成本高、不易收敛等情况, 因此提出了辅助神经网络, 就是在这个神经网络外, 再搭建小的神经网络, 来辅助计算梯度
 - 大致是这个意思
- 这个辅助神经网络概念引入后, 我们就可以做更细粒度的事情了, 同时网络形式上也往全连接层网络靠也就是 $W = x_11W_11 + x_12W_12 + x_13W_13 + x_21W_21 + x_22W_22 + x_23W_23 + \dots$ 整体表现形式大概就是这样子, 我们做全连接层, 每一层训练的参数都是 x_{n1}, x_{n2}, x_{n3} 这样的 $[1,3]$ 维的向量, 更进一步呢, 我们可以分qv矩阵, 将 $W_n x$ 拆成 $q_n x$ 和 $v_n x$, 再进一步就是 A_{qn1} 和 A_{qn2} 这种, 也可以把我们的网络复杂起来
- 我们的贡献
 - 基于Boosting算法, 提出了一种新的lora合并方式
 - 将辅助网络引进, 来改善lora合并表现
 - 提供常用任务的lora“预设”参数