# Bert

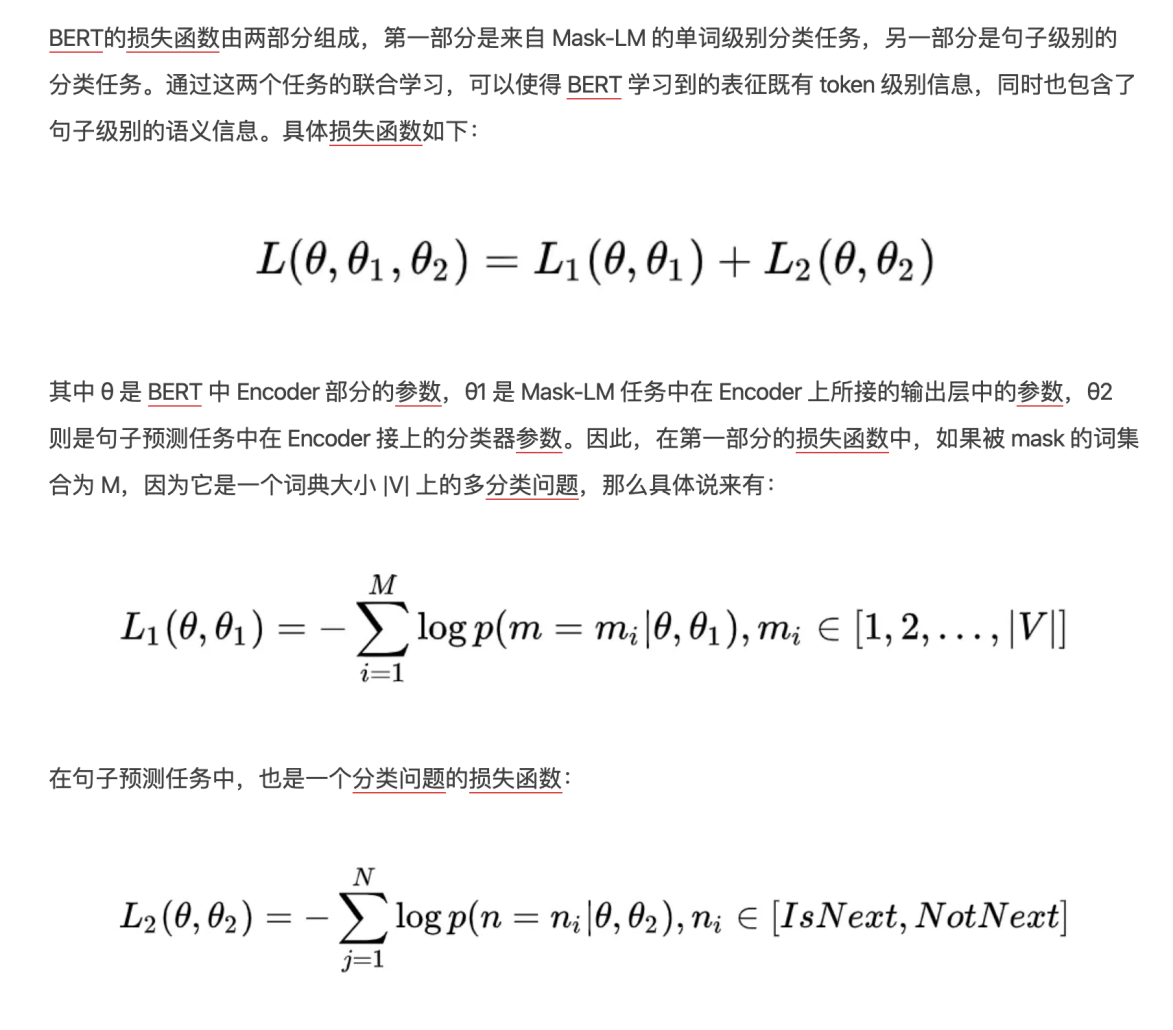
知道了transformer之后再理解bert就简单多了。

bert思想：

1. 将下游的完形填空和下一个句子预测放在了训练词向量的任务中

bert损失函数：

预测mask值的损失函数和预测下一个句子的二分类的损失函数之和



bert怎样做预训练

这个博客上内容非常多了。不写了。

bert怎样做微调

