# Bert

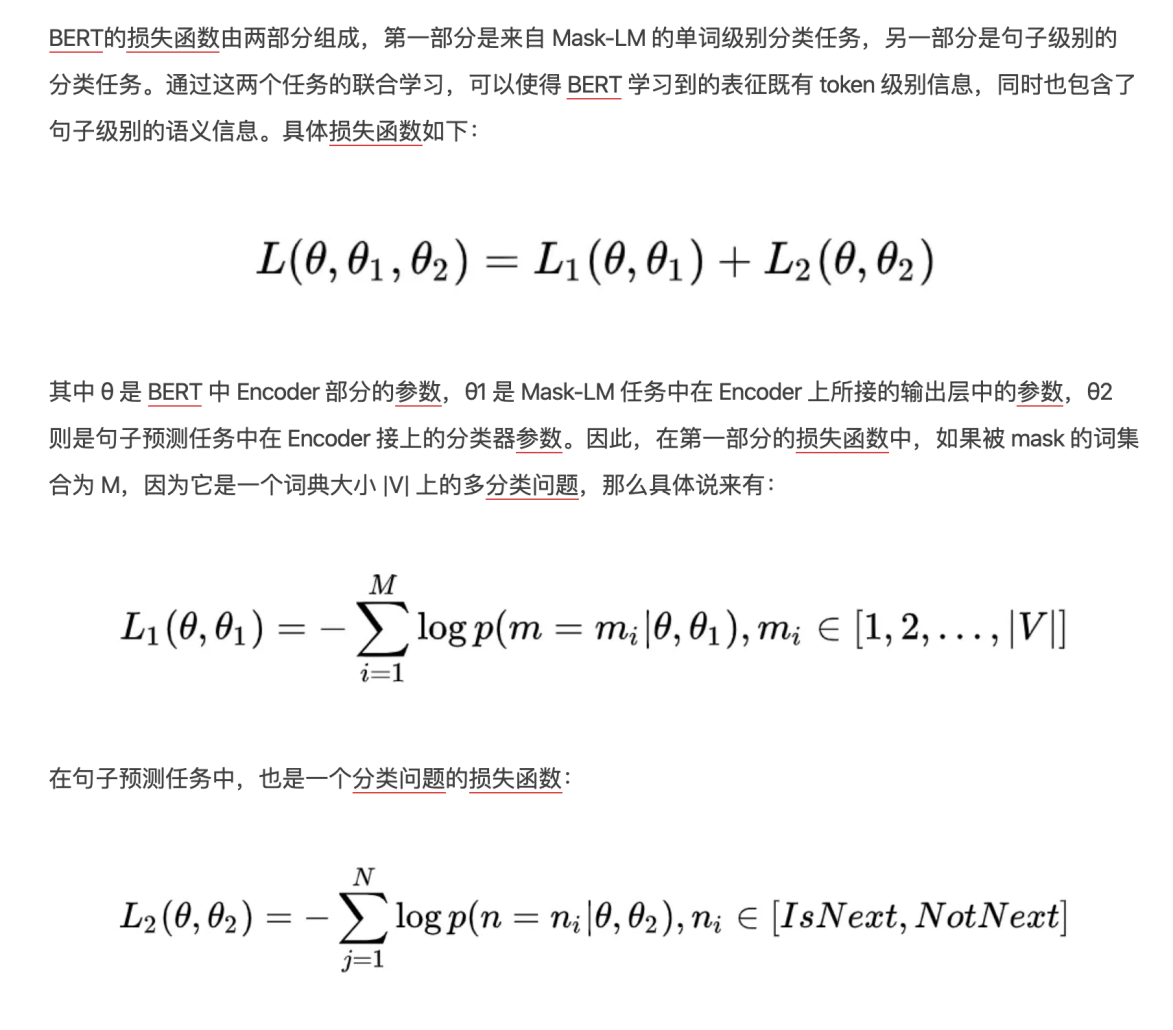
知道了transformer之后再理解bert就简单多了。

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

这个问题比较关键。

首先，看你的任务是什么？

1. 文本分类

文本分类任务我们需要的是文本的向量表示，这个时候并没有下一个句子预测的思想，因此，在微调的时候必须mask词，当成训练一样去调整词向量和文档向量

1. 标注

譬如分词等任务。这个时候语料往往是多篇文档之类的东西。这个时候便有了当前这句话和下一句话的概念。考虑损失函数，因此可以不mask，直接得到词向量。

1. 问答匹配

问答匹配，是有下一个句子的概念的。同上。

bert怎样做实时预测

字作为输入，不需要mask，代入模型，模型的输出就是词的向量表示和句子的向量表示，不需要更新模型。

