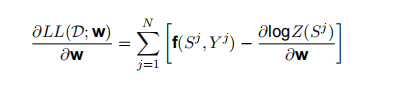
从模型定义来看，CRF和线性模型有很多相似之处，似然函数的转化和最大熵模型差不多，都是取log将除法变成减法。



对减去的后项

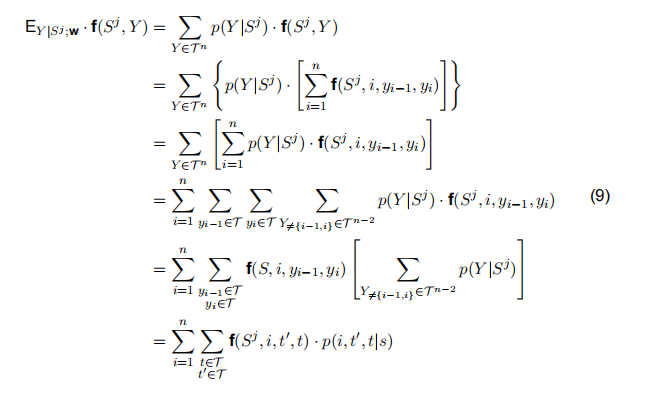


求偏导转化后可以变为

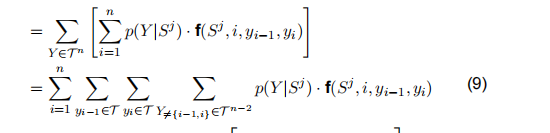


Y’我看作词性序列。那么这式子就代表句子S标为各个词性序列Y’的可能性与对应抽取的特征的点积，它们的累加之和。相当于所有可能的序列的得分加起来。

继续推导：



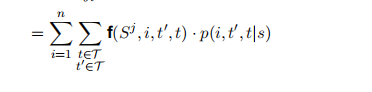
前面几步都好理解，



这一步有点难懂，四个累加太抽象了。

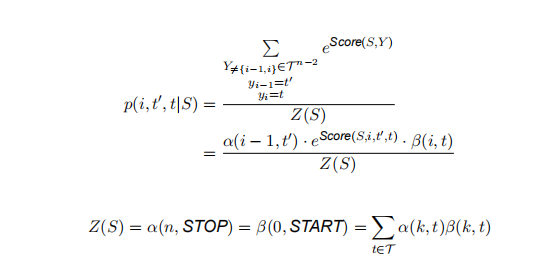
我的理解是，这个转换大致意义为给词性序列所有词性打分，乘上句子Sj表现为各个序列Y的概率，里面这些Y序列，是不考虑i-1和i号词以后，其他所有可能的序列。

等于和前面的换了一下计算顺序，不先给序列打分，每个词性单独先算。



最终得到的就是这个结果，我理解为句子s第i-1号词的词性为t’，i号为t的概率乘不加权重的分值。这一整套转化就是把原来一起算的序列拆开了。

前向后向算法，属于动态规划，HMM里也有类似算法，这部分比较好理解，前向算法计算前一个词为各种词性时当前词为词性t的得分之和。后向类似，只是从后往前传播。



这里，我的理解是，k处词性为t的概率为a（k，t）\*β（k，t），除去i-1和i号词后，用a（i-1，t’），β（i，t）计算前面和后面的概率和，乘i-1位置为t’，i位置为t的得分，这样就解决了Y’序列组合数极多的问题，有31^N种组合，一个个算复杂度非常高。