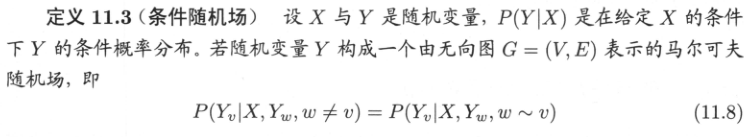
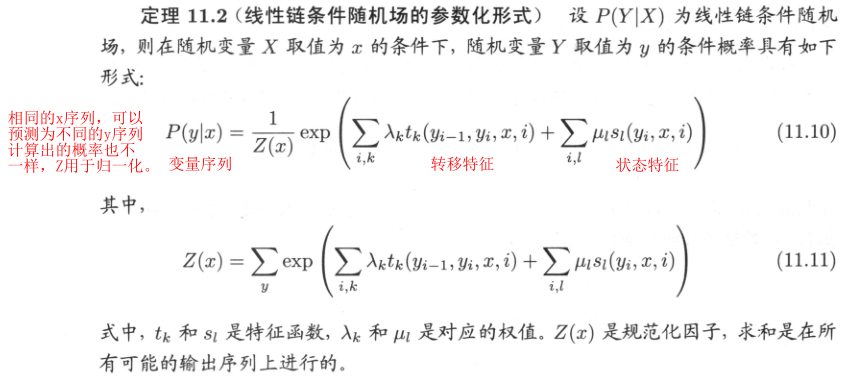
### 11.1.2 概率无向图模型的因子分解

将概率无向图模型的联合概率分布表示为最大团上的随机变量函数乘积形式的操作，称为概率无向图模型的因子分解。  
11.2 条件随机场的定义



### **11.2.2 条件随机场的参数化形式**



理解：求已知观测序列x的条件下，状态(标记)序列为y的概率：

1、主题部分为特征函数值的计算

2、以转移特征为例，在每一个时间步，可以计算k个特征值

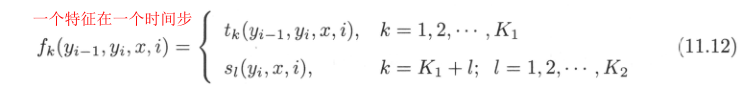
3、对特征值加权求和，在求其指数并归一化即为该条件概率

### 11.2.3 向量形式(时间维求和)

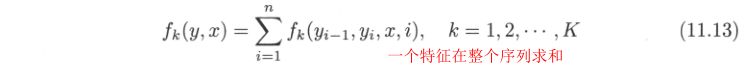
将参数形式的权重和特征函数统一符号描述，为此有：

1、假设K1个转移特征，K2个状态特征，令K=K1+K2

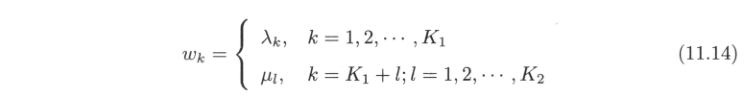
2、统一的特征描述为：



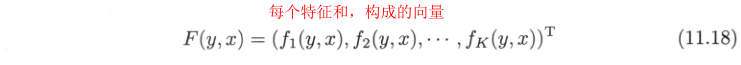
3、一个特征在所有的时间步的斩获(数量值)



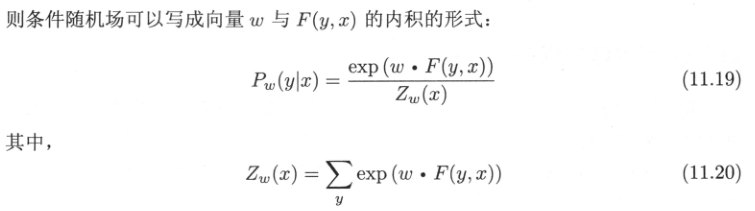
4、权值统一化：



5、特征向量(每个元素，表示一个特征在整个序列上的斩获)



6、内积形式

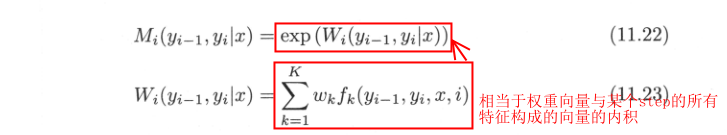


参数形式：求和

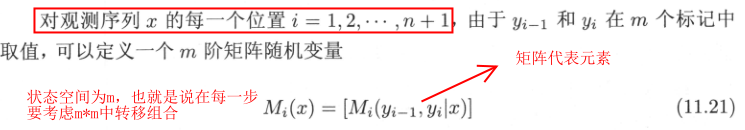
内积形式：内积(等价于求和)

### 11.2.4 矩阵形式(特征维求和)

1、单步特征和：单个时间步基于所有特征的斩获



2、单步矩阵：



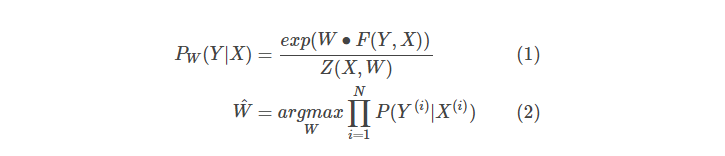
为什么是m阶呢？因为考虑不同的标记组合带入到M(x)中。基于此可以求所有y序列的条件概率

## 11.4 参数学习算法

input: (xi, yi) N对，最大似然函数

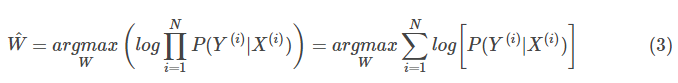
output: 参数λ(特征权重),η(状态权重)

向量形式：

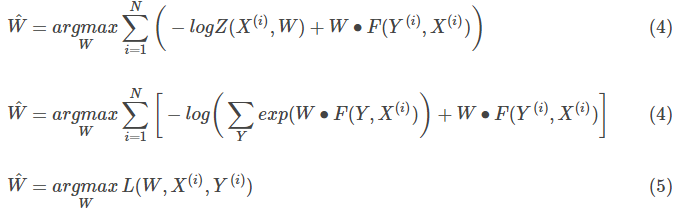


其中，F为K(特征总数)维特征向量，W为K为权重，目标就是最大似然。

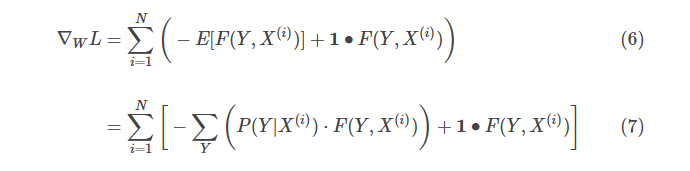
1、对数似然为：



2、带入条件概率，并将归一化因子展开为：



3、梯度计算



由于F为特征函数，第二项可以直接求解。现在考虑第一项z(x)：

将特征按照时间步展开：注意此时f为单步K维特征向量

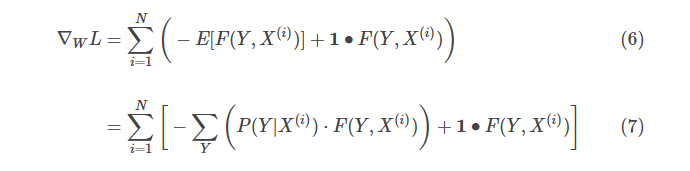
将时间步前提：

将Y拆分：

将pdf通过"积分"化为"边缘概率密度"

由于p可以由前后向算法计算得出，f可以直接计算，至此已经没有变量。

综上：



其中：F(Y,X)=∑tf(yt-1, yt, Xi)

参数更新：

W=step\*

注：求充分统计量的均值用到了：log-partition Function,具体理论来源于指数族分布相关知识。

参考1：<https://www.bilibili.com/video/BV19t411R7QU>

参考：2 <https://anxiang1836.github.io/2019/11/05/NLP_From_HMM_to_CRF/#%E7%BA%BF%E6%80%A7%E9%93%BE%E6%9D%A1%E4%BB%B6%E9%9A%8F%E6%9C%BA%E5%9C%BA-Linear-CRF>