**Negative Log-Likelihood Loss（负log likelihood损失函数）**

推导过程如下

假定我们有P个样本构成的语料集合S：[{}，{},…{}]。

我们希望产生这个S的概率是最大的，是最自然的，而不是别别扭扭产生的。

假定每个样本产生是独立的，则有下面公式

(1)

对公式（1）求负对数得到公式（2）

(2)

由Gibbs distribution公式（3）

(3)

(3)带入(2)得到

(4)

在公式（4）上除P和B，因为除的是常数，对最终求导的方向没有影响。 得到公式(5)

(5)

(5)是全部样本梯度，如果只对单个样本求梯度，有公式(6)

(6)

这里假设

(7)

公式7中应用了一次Gibbs distribution公式如下

这样公式（6）可以写成公式（8）

(8)

可以理解为由{Xi,Yi}这个样本产生的对的梯度，等于{Xi,Yi} 产生的能量对的偏导减去该模型对Xi 产生的所有y上的偏导的期望。前项往往比较好求，后者用Monte-Carlo方法求