序列模型

在时间 t 观察到 x_t，得到 T 个不独立的随机变量
$(x_1, \ldots, x_T) \sim p(X)$，

条件概率

$p(a, b) = p(a)\, p(b \mid a) = p(b)\, p(a \mid b)$

联合分布的链式展开

当前的数据只和过去的数据点相关。

正序展开：

$$p(x_1, \ldots, x_T) = \prod_{t=1}^{T} p(x_t \mid x_1, \ldots, x_{t-1}), \quad \text{其中约定 } p(x_1 \mid \cdot) = p(x_1)$$

反序展开：

$$p(x_1, \ldots, x_T) = \prod_{t=T}^{1} p(x_t \mid x_{t+1}, \ldots, x_T), \quad \text{其中约定 } p(x_T \mid \cdot) = p(x_T)$$

根据已知的数据进行建模，也称为**自回归模型**。

马尔可夫假设

假设当前数据只跟 τ 个过去数据点有关：

$$p(x_t \mid x_1, \ldots, x_{t-1}) \approx p(x_t \mid x_{t-\tau}, \ldots, x_{t-1})$$

潜变量模型

引入潜变量h_t来表示过去的信息$h_t = f(x_1, \ldots, x_{t-1})$
这样的话$x_t = p(x_t|h_t)$

时序模型汇总当前的数据跟之前观察到的数据有关
自回归模型使用自身过去的数据来预测未来
马尔科夫模型假设当前只和最近少数数据相关从而简化模型
潜变量模型使用潜变量来概括历史信息

语言模型

给定文本序列$x_1, \ldots, x_T$，语言模型的目标是估计联合概率$p(x_1, \ldots, x_T)$，预测整个文本生成概率

应用包含
作预训练模型
生成文本给定前面的词不断向后预测后续的文本
判断哪个单词序列更常见

使用计数来建模
$p(x, x') = p(x)p(x'|x) = \frac{n(x)}{n} \times \frac{n(x,x')}{n(x)}$
n总词数，$n(x)$，$n(x, x')$是单个单词和连续单词对的出现次数

当序列很长的时候因为文本量不够大，使用马尔科夫假设来缓解这种问题
一元语法：$p(x_1, x_2, x_3, x) = p(x_1)p(x_2)p(x_3)p(x)$
二元语法：$p(x_1, x_2, x_3, x) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x|x_3)$
三元语法：$p(x_1, x_2, x_3, x) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x|x_2, x_3)$