序列模型

动手学深度学习 v2

李沐 · AWS

序列数据

- 实际中很多数据是有时序结构的
- 电影的评价随时间变化而变化
  - 拿奖后评分上升，直到奖项被忘记
  - 看了很多好电影后，人们的期望变高
  - 季节性：贺岁片、暑期档
  - 导演、演员的负面报道导致评分变低

序列数据 - 更多例子

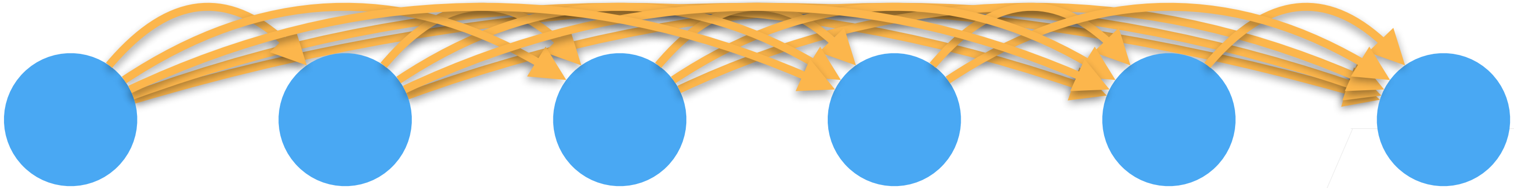
- 音乐、语言、文本、和视频都是连续的
  - 标题“狗咬人”远没有“人咬狗”那么令人惊讶
- 大地震发生后，很可能会有几次较小的余震
- 人的互动是连续的，从网上吵架可以看出
- 预测明天的股价要比填补昨天遗失的股价的更困难

统计工具

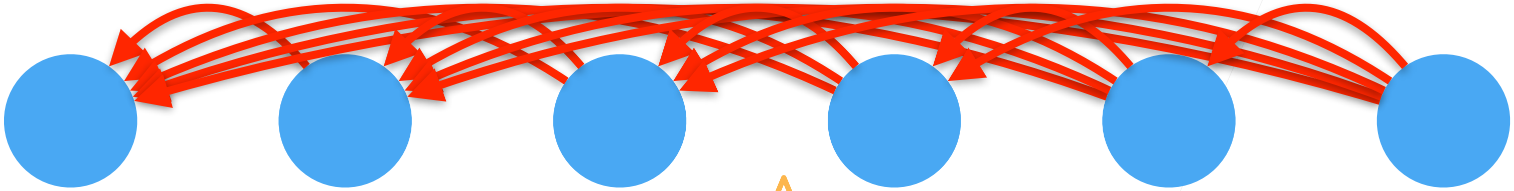
- 在时间 t 观察到 x_t，那么得到 T 个不独立的随机变量
  $(x_1, \ldots x_T) \sim p(\mathbf{x})$
- 使用条件概率展开
  $p(a, b) = p(a)p(b \mid a) = p(b)p(a \mid b)$

统计工具

$$p(\mathbf{x}) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_1, x_2) \cdot \ldots p(x_T \mid x_1, \ldots x_{T-1})$$



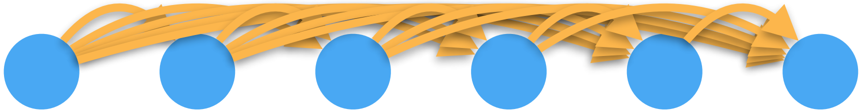
$$p(\mathbf{x}) = p(x_T) \cdot p(x_{T-1} \mid x_T) \cdot p(x_{T-2} \mid x_{T-1}, x_T) \cdot \ldots p(x_1 \mid x_2, \ldots x_T)$$



物理上不一定可行

序列模型

$$p(\mathbf{x}) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_1, x_2) \cdot \ldots p(x_T \mid x_1, \ldots x_{T-1})$$



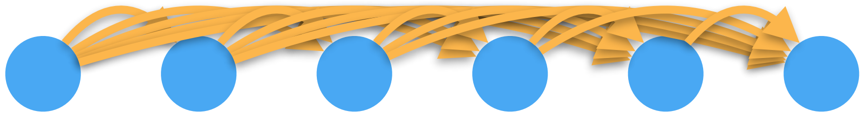
- 对条件概率建模

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

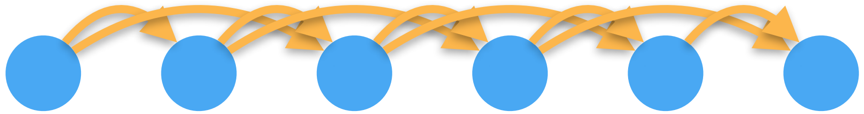
对见过的数据建模，也称自回归模型

方案 A - 马尔科夫假设

$$p(\mathbf{x}) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_1, x_2) \cdot \ldots p(x_T \mid x_1, \ldots x_{T-1})$$



- 假设当前当前数据只跟 τ 个过去数据点相关

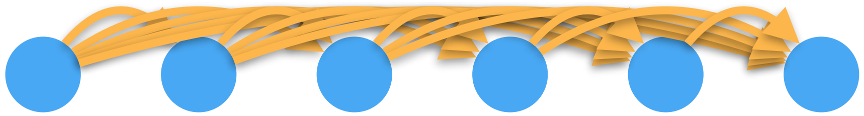


$$p(x_t \mid x_1, \ldots x_{t-1}) = p(x_t \mid x_{t-\tau}, \ldots x_{t-1}) = p(x_t \mid f(x_{t-\tau}, \ldots x_{t-1}))$$

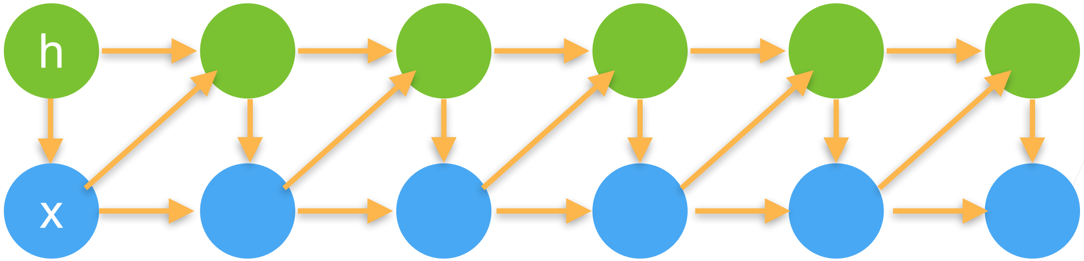
例如在过去数据上训练
一个MLP模型

方案 B - 潜变量模型

$$p(\mathbf{x}) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_1, x_2) \cdot \ldots p(x_T \mid x_1, \ldots x_{T-1})$$



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



总结

- 时序模型中，当前数据跟之前观察到的数据相关
- 自回归模型使用自身过去数据来预测未来
- 马尔科夫模型假设当前只跟最近少数数据相关，从而简化模型
- 潜变量模型使用潜变量来概括历史信息