

# 15. 表示学习 + 16. 结构化概率模型

- 任务：
  - 15+16: ddl:10/10
    - ☐ 更新Pytorch代码
    - ☐ 优化和更新本章内容
      - ☐ 核对原书中文翻译
      - ☐ 撰写笔记
    - ☐ 提交Pull Request
  - 其他可优化的地方：
    - 可将笔记合并为tex格式，以便生成电子书
    - 修正其他章节内容和.md格式、排版问题

## 15. 表示学习

$$\nabla_{\mathbf{c}} L = \sum_t \left( \frac{\partial o^{(t)}}{\partial \mathbf{c}} \right)^T \nabla_{\mathbf{o}^{(t)}} L$$

$$\nabla_{\mathbf{b}} L = \sum_t \left( \frac{\partial h^{(t)}}{\partial \mathbf{b}} \right)^T \nabla_{\mathbf{h}^{(t)}} L$$

$$\nabla_{\mathbf{V}} L = \sum_t \sum_i \left( \frac{\partial L}{\partial o_i^{(t)}} \right)^T \nabla_{\mathbf{o}^{(t)}} L h^{(t-1)^T}$$

$$\nabla_{\mathbf{W}} L = \sum_t \sum_i \left( \frac{\partial L}{\partial h_i^{(t)}} \right)^T \nabla_{\mathbf{h}^{(t)}} L x^{(t-1)^T}$$

$$\nabla_{\mathbf{U}} L = \sum_t \sum_i \left( \frac{\partial L}{\partial h_i^{(t)}} \right)^T \nabla_{\mathbf{h}^{(t)}} L x^{(t)^T}$$

除了梯度表达式不同，RNN的反向传播算法和DNN区别不大。