

## 神经网络

Lecturer: Changshui Zhang      zcs@mail.tsinghua.edu.cn

Student: \_\_\_\_\_

本次作业我们将完成 CNN 和 RNN 的 BP 算法推导。

1. 考虑表 1 中的简单 CNN 网络，其中卷积层、全连接层、输出层的非线性单元均为  $f$  函数，输出层损失函数为  $L$ 。该网络中需要学习的参数有卷积核权值  $K$  ( $3 \times 3$  大小)、卷积核偏置  $B$  (单值)、第一全连接层权值  $w1$  (大小  $4 \times 2$ ) 及偏置  $b1$  (大小  $1 \times 2$ )，第二全连接层权值  $w2$  (大小  $2 \times 1$ ) 及偏置  $b2$  (单值)。请推导该 CNN 结构的 BP 算法，给出卷积核权值  $K$  及偏置  $B$  从  $t$  步到  $t+1$  步的更新公式。

表 1. 简单 CNN 结构

Input	(4,4)
Conv Kernel (no padding)	(3,3)
Conv output	(2,2)
Flatten	(4,)
Dense	(2,)
Output	1

注意：合理运用符号及上下标对网络结构中的其他相关变量进行自行命名，便于公式表达。

提示：建议从全连接神经网络的 BP 算法中总结梯度的传播规律，推广至 CNN。

可参考 Notes on Convolutional Neural Networks:

[http://cogprints.org/5869/1/cnn\\_tutorial.pdf](http://cogprints.org/5869/1/cnn_tutorial.pdf)

2. 某 RNN 结构，时间序列长度  $T = 3$ ， $t$  时刻隐层为  $H_t$ ，输出为  $Y_t$ ，输入为  $X_t$ ，前向计算如下：

$$H_t = \text{sigmoid}(UX_t + WH_{t-1})$$

$$Y_t = \text{softmax}(VH_t)$$

其中  $X_t \in R^{5 \times 1}$ ,  $H_t \in R^{4 \times 1}$ ,  $Y_t \in R^{3 \times 1}$ ,  $W \in R^{4 \times 4}$ ,  $U \in R^{4 \times 5}$ ,  $V \in R^{3 \times 4}$

1) 采用 cross-entropy 损失，RNN 网络优化的目标损失函数表达式

2) 采用 SGD 优化，学习率为  $lr$ ，推导对某个样本  $X \in R^{5 \times 3}$ ，标签  $Y \in R^{3 \times 3}$ ，参数矩阵  $W$ ， $U$ ， $V$  的更新公式。（详细写出链式法则推导过程）

本次作业没有编程作业，希望同学们花时间熟悉相关深度学习平台的使用：

建议掌握：

1) Python 的基本编程；

2) 底层库：tensorflow、pytorch 等任选一个；

3) 基于底层库的一些现成的工具包：keras、lasagne。