

# 计算机视觉课程第二次作业：softmax与相关梯度推导

## 作业内容：

$\text{softmax}$ 函数是计算机视觉中常用到的一种指数规范化函数， $\text{softmax}$ 可以将向量的和规范化为1，因此常用来将一组特征映射成一组概率分布。 $\text{softmax}$ 函数的定义如下：

对于向量 $\mathbf{z} = [z_1, z_2, \dots, z_k]^T \in \mathbb{R}^k$ ，输出 $\mathbf{y} = \text{softmax}(\mathbf{z}) \in \mathbb{R}^k$ ，其中第 $i$ 个元素为：

$$y_i = \frac{\exp(z_i)}{\sum_{j=1}^k \exp(z_j)}$$

1. 在图像处理或深度学习的工具箱中，通常会利用 $\text{softmax}$ 的平移不变性将 $\text{softmax}(\mathbf{z})$ 的运算等价地转换为 $\text{softmax}(\mathbf{z} - \max(\mathbf{z}))$ ，利用这一技巧，可以有效地避免数值溢出。请证明 $\text{softmax}(\mathbf{z})$ 的平移不变性，即证明： $\text{softmax}(\mathbf{z}) = \text{softmax}(\mathbf{z} + \mathbf{c})$ ，其中： $\mathbf{c} = [c, c, \dots, c]^T \in \mathbb{R}^k$ 。（30分）

2. 令 $y_i = \text{softmax}(\mathbf{z})_i$ ， $1 \leq i \leq k$ 。请计算 $z_i$ 的梯度 $\frac{\partial y}{\partial z_i}$ 。

提示： $\text{softmax}$ 运算中每项元素 $y_i (i = 1, \dots, k)$ 的分母都是所有输入元素 $z_j (j = 1, \dots, k)$ 的指数幂的和，因此元素 $z_j$ 参与了所有元素 $y_i$ 的规范化过程。那么，在计算元素 $z_j$ 的梯度时应该考虑所有元素 $y_i$ 对其的影响。（30分）

3. 请进一步讨论在CNN中使用 $softmax$ 的情况。在CNN中，卷积层、正规化层、非线性激活层构成了CNN的一个基本单元，多个基本单元的堆叠就形成了CNN的主体。为了将预测值与真值比较、计算损失函数，在CNN的末端还会再加上全连接层、 $softmax$ 与损失函数。现将这一网络结构简化：如下图所示，特征 $x$ 依次进入线性层、ReLU层、线性层、 $softmax$ 层、损失函数层。

$$p = W_1^T x \mapsto r = \text{ReLU}(p) \mapsto z = W_2^T r \mapsto y = \text{softmax}(z) \mapsto l = \text{loss}(y)$$

其中： $x \in \mathbb{R}^d$   $p, r \in \mathbb{R}^{k_1}$   $z, y \in \mathbb{R}^{k_2}$   $l \in \mathbb{R}$   $W_1 \in \mathbb{R}^{d \times k_1}$   $W_2 \in \mathbb{R}^{k_1 \times k_2}$ .

请根据所给的模块结构，模拟推导CNN中各层后项传播的过程，做如下推导：（40分）

(1) 已知 $\frac{\partial l}{\partial y}$ ，推导 $\frac{\partial l}{\partial z}$ .

(2) 已知 $\frac{\partial l}{\partial z}$ ，推导 $\frac{\partial l}{\partial r}$ .

(3) 已知 $\frac{\partial l}{\partial r}$ ，推导 $\frac{\partial l}{\partial p}$ .

(3) 已知 $\frac{\partial l}{\partial p}$ ，推导 $\frac{\partial l}{\partial x}$ .

## 作业要求：

- 要有完整的推导过程，直接列出结果则不得分。
- 作业中出现的公式需用公式编辑器编辑，不可以直接以文本形式出现。

## 作业提交：

- 作业请提交pdf格式，pdf文件命名方式为：第二次作业\_学号\_姓名.pdf，如：第二次作业\_21709001\_李小明.pdf。
- 第二次作业、第三次作业请一并提交，以两个附件发送到 [gao\\_zilin@126.com](mailto:gao_zilin@126.com) 邮箱中。其中，第二次作业的pdf为一个附件，第三次作业的压缩包为另一个附件。邮件主题命名方式：第二三次作业\_学号\_姓名。
- 如果作业提交后想要再提交新的版本，请重新发送第二三次的所有作业，并在上述1.&2.涉及到的所有命名方式后加上版本号，如：第二次作业\_21909001\_李晓明\_v2.pdf。评分时将以截止时间前的最后一次提交版本为准。
- 第二、三次作业提交的截止时间：6月10日23:59，逾期视作未提交。