

作业 3

2022 年 3 月 30 日

本次作业通过卷积神经网络实现非线性分类任务，帮助同学们理解卷积神经网络的基本原理并掌握卷积神经网络的实现方法。作业内容为卷积神经网络的实现及模型优化求解方法，并在数据集上进行训练、测试和可视化。具体任务分为理论部分、编程部分以及作业报告。其中理论部分包含第 1, 2 题，所有同学均需完成，答案附在作业报告中；编程部分包含第 3、4 题，作业报告为第 5 题，已确认自选课题的同学需完成第 6 题。

1. 单选题 (15 分)
2. 计算题 (15 分)
3. 完成卷积神经网络分类任务的程序代码 (30 分)
4. 训练/测试/可视化 (30 分)
5. 撰写作业报告 (10 分)
6. 汇报自选课题进度 (70 分) *

理论部分

1 单选题 (15 分)

1.1 以下关于卷积神经网络的说法错误的是

- (A) 全卷积网络常用于图像分割任务。
- (B) Batch Normalization 和残差连接均可用于缓解模型训练过程中梯度消失的问题。
- (C) 可变形卷积中的卷积核位移矢量需要用另一个网络来学习。
- (D) 只有通过增加卷积核的大小才能增大感受野。

- 1.2 考虑某卷积层，已知输入数据的尺寸 (s_in)，卷积核的尺寸 ($kernel_size$)，步长 ($stride$) 和边界延拓 ($padding$)。以下关于输出数据的尺寸 (s_out) 的计算正确的是：

注： $\lfloor \cdot \rfloor$ 表示向下取整， $\lceil \cdot \rceil$ 表示向上取整

- (A) $s_out = \lfloor (s_in - kernel_size + padding) / stride \rfloor + 1$
(B) $s_out = \lfloor (s_in - kernel_size + 2 * padding) / stride \rfloor + 1$
(C) $s_out = \lceil (s_in - kernel_size + 2 * padding) / stride \rceil + 1$
(D) $s_out = \lceil (s_in - kernel_size + padding) / stride \rceil + 1$

- 1.3 考虑某 batch normalization 层，其通道数为 1，在训练过程中计算得到的均值和方差分别为 -0.3 和 0.16。在测试过程中输入批次包含两个样本，输入该层的特征为 0.1 和 -0.5，那么该层的输出特征为：

注：batch normalization 计算公式为 $y = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$ ，本题中假设 $\gamma = 1, \epsilon = 0, \beta = 0$

- (A) 0.3, -0.3
(B) 1, -1
(C) 1, -0.5
(D) 2.5, -1.25

- 1.4 考虑某最大池化层，其 $kernel_size$ 为 2，步长 ($stride$) 为 2，边界延拓 ($padding$) 为 1。假设输入为 X ，输出为 Y 。现已知：上下左右都有延拓

$$X = \begin{bmatrix} 0.1 & -0.2 & -0.1 & -0.2 \\ 0.0 & 0.3 & 0.2 & -0.3 \end{bmatrix}$$

那么 Y 的值为：

- (A) $Y = [0.3, 0.2]$
(B) $Y = [0.1, 0.3, 0.2]$

$$(C) \quad Y = \begin{bmatrix} 0.1 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \end{bmatrix}$$

$$(D) \quad Y = \begin{bmatrix} 0.1 & -0.1 & -0.2 \\ 0.0 & 0.3 & -0.3 \end{bmatrix}$$

1.5 对于上一题中的最大池化层，假设损失函数为 L ，反向传播的梯度为 $\frac{\partial L}{\partial X}, \frac{\partial L}{\partial Y}$ 。现已知：

$$\frac{\partial L}{\partial Y} = \begin{bmatrix} -0.1 & 0.2 & 0.1 \\ 0.2 & -0.3 & -0.1 \end{bmatrix}$$

那么 $\frac{\partial L}{\partial X}$ 的值为：

$$(A) \quad \frac{\partial L}{\partial X} = \begin{bmatrix} -0.1 & 0.0 & 0.2 & 0.1 \\ 0.2 & -0.3 & 0.0 & -0.1 \end{bmatrix}$$

$$(B) \quad \frac{\partial L}{\partial X} = \begin{bmatrix} -0.1 & 0.0 & 0.0 & 0.0 \\ 0.2 & -0.3 & 0.0 & 0.0 \end{bmatrix}$$

$$(C) \quad \frac{\partial L}{\partial X} = \begin{bmatrix} -0.1 & 0.0 & 0.1 & 0.0 \\ 0.2 & -0.3 & 0.2 & -0.1 \end{bmatrix}$$

$$(D) \quad \frac{\partial L}{\partial X} = \begin{bmatrix} -0.1 & 0.2 & 0.2 & 0.1 \\ 0.2 & -0.3 & -0.3 & -0.1 \end{bmatrix}$$

2 计算题 (15 分)

2.1 已知某卷积层的输入为 X (该批量中样本数目为 1，输入样本通道数为 1)，采用一个卷积核 W ，即卷积输出通道数为 1，卷积核尺寸为 2×2 ，卷积的步长为 1，无边界延拓，偏置量为 b ：

$$X = \begin{bmatrix} -0.5 & 0.2 & 0.3 \\ 0.6 & -0.4 & 0.1 \\ 0.4 & 0.5 & -0.2 \end{bmatrix}, W = \begin{bmatrix} -0.2 & 0.1 \\ 0.4 & -0.3 \end{bmatrix}, b = 0.05$$

2.1.1 请计算卷积层的输出 Y 。

2.1.2 若训练过程中的目标函数为 L ，且已知 $\frac{\partial L}{\partial Y} = \begin{bmatrix} 0.1 & -0.2 \\ 0.2 & 0.3 \end{bmatrix}$ ，请计算 $\frac{\partial L}{\partial X}$ 。

注：本题的计算方式不限，但需要提供计算过程以及各步骤的结果。

编程部分

编程部分包括第 3，4 题。

3 完成卷积神经网络分类任务的程序代码（30 分）

使用卷积神经网络机完成识别英文字符图像（非线性分类）任务，在本任务中，我们将使用卷积神经网络识别英文字符图像，即输入一张图像，模型输出识别结果。请注意图像样本包括大小写英文字母，识别时不区分大小写形式，即大写字母“A”和小写字母“a”都对应于一个类别，输出时转为大写字母“A”显示。

程序清单如下：

文件或目录	说明	注意事项
hw3.zip	作业 2 程序压缩包	解压可以得到下列文件
\data	存放本次作业所用数据集	请勿修改
\visual	存放本次作业的可视化结果	请勿修改
conv.py	卷积层定义	需要完成代码
pool.py	最大池化层定义	已完成代码
batchnorm.py	Batch Normalization 定义	已完成代码
network.py	卷积神经网络结构定义	需要完成代码
train.py	训练程序	需要完成代码
test.py	测试程序	已完成代码
visualize.py	可视化程序	已完成代码
cal_entropy.py	熵计算程序	已完成代码

需要完成的代码清单如下：每处需要完成的地方都有代码提示和步骤提示
conv.py 文件中待完成内容：

序号	行号	内容	说明
TODO 1.1	46	完成卷积层输出特征图尺寸计算	注意尺寸的数值必须为整数
TODO 1.2	49	完成卷积层的前向计算	请按照 step 1-3 完成卷积层的前向计算
TODO 1.3	94	完成卷积层的反向计算	请按照 step 1-4 完成卷积层的反向计算
TODO 1.4	168	完成卷积层参数的初始化	注意参数各维度的大小并设置参数为可训练

network.py 文件中待完成内容：

序号	行号	内容	说明
TODO 2.1	35	完成卷积神经网络中卷积层和池化层的定义	请按照注释中提供的网络结构和参数完成定义
TODO 2.2	67	完成卷积神经网络中线性层的定义	请按照注释中提供的网络结构和参数完成定义
TODO 2.3	106	完成卷积神经网络的计算	请按照 step 1-3 完成卷积神经网络的计算

train.py 文件中待完成内容：

序号	行号	内容	说明
TODO 3.1	4	使用 ImageFolder 构建数据集	按照 ImageFolder 的定义来使用

4 训练/测试/可视化 (30 分)

(1) 在默认参数下，没有启用 batch normalization，dropout 的概率为 0.0。同学们选择启用 batch normalization 或者选择不同的 dropout 概率来观察和记录其对训练、验证和测试性能的影响。其中训练和测试的运行命令示例如下：

注意：建议使用 ckpt_path 参数指定模型参数存放位置，便于测试和可视化；对于不同参数对结果的影响，最好能进行分析；训练过程中 loss 和验

证集上 accuracy 的变化会在训练完成后以图片形式显示，并自动保存在 ckpt_path 文件夹中，也可以选择手动保存。

1) 训练模型，使用默认配置的命令如下：

```
python train.py --ckpt_path ckpt
```

2) 训练模型，启用 batch normalization 的命令如下：

```
python train.py --ckpt_path bn_ckpt --bn
```

3) 训练模型，改变 dropout 概率为 0.3 的命令如下：

```
python train.py --ckpt_path drop_ckpt --dropout 0.3
```

4) 测试默认参数模型命令如下：

```
python test.py --ckpt_path ckpt
```

(2) 可视化：任意选取一个模型，使用 visualize.py 对卷积核和特征图进行可视化，并利用 t-SNE 将其将线性层高维特征降维到二维空间。将使用模型的情况和可视化结果附在实验报告中，**最好能简要描述观察到的结果**。可视化的运行命令示例如下：

注意：实验提供的卷积神经网络包含 5 个卷积层，可以任意选择若干个进行可视化；在特征图可视化时，可以选择任意图片编号；可视化结果会自动保存在 visual 文件夹中，也可以选择手动保存图片。

1) 可视化第 0 层卷积层的卷积核：

```
python visualize.py --type filter --ckpt_path bn_ckpt --layer_idx 0
```

2) 可视化第 100 张图像第 1 层卷积层的输出特征图：

```
python visualize.py --type feature --ckpt_path bn_ckpt --layer_idx 1 --image_idx 100
```

3) t-SNE 显示分类结果：

```
python visualize.py --type tsne --ckpt_path bn_ckpt --layer_idx 1
```

(3) 熵计算：请选择你认为效果最好的一个模型，使用 cal_entropy.py，先计算任意图像的图像熵、特征熵以及模型输出预测概率的信息熵（参考第一周课件第 62 页），然后计算整个测试集以上熵的平均值。请把使用模型的情况和熵的计算结果写入作业报告。熵计算的运行命令如下：

1) 单张图像熵计算：

```
python cal_entropy.py --mode single --ckpt_path bn_ckpt --im_path data/test/B/1390.jpg
```

2) 测试集熵计算：

```
python cal_entropy.py --mode dataset --ckpt_path bn_ckpt
```

5 撰写作业报告（10 分）

将 hw3 目录和作业报告打包为一个文件（例如 *.zip）提交到网络学堂。作业报告中包括选择题答案，计算题的解题步骤及答案、任务 3、4 运行结果

及分析，本次作业遇到的问题及解决方法，对本次作业的意见及建议。推荐同学们使用随作业发布的 LaTeX 模板 HW3-template.zip 完成作业报告。

6 自选课题进度汇报（70 分）*

请已确认自选课题的同学，完成简短的自选课题工作进度汇报，例如，文献阅读、或者研究方案设计、或者原型系统搭建及实验结果等内容。

关于作业迟交的说明：由于平时作业计入总评成绩，希望同学们能按时提交作业。若有特殊原因不能按时提交，请在提交截止时间之前给本次作业责任助教发 Email 说明情况并给出预计提交作业的时间。对于未能按时说明原因的迟交作业，将酌情扣分。

本次作业责任助教为姚刚 (Email: yg19@mails.tsinghua.edu.cn)。

7 附录

我们通过 argparse 库进行参数设置，可以查看 train.py, test.py, visualize.py, cal_entropy.py 中可以调节的参数，可调节参数如下表所示。

	参数	说明
共有参数	ckpt_path	保存训练模型的路径
	device	代码运行时使用的设备, 可以为 cpu 或 cuda
train.py	train_file_path	存放训练图像的文件夹路径, 不用修改
	val_file_path	存放验证图像的文件夹路径, 不用修改
	epoch	训练轮数, 默认 10
	batch_size	训练时每一批数据中包含的数据个数
	lr	学习率
	momentum	SGD 优化器的动量, 只有 optim_type 设为 sgd 时使用
	weight_decay	L2 正则化系数
	optim_type	优化器类型, 可以设为 sgd, adam, adagrad, rmsprop 或 adadelta
	bn	是否使用 batch normalization
	dropout	dropout 概率
	max_ckpt_save_num	保存训练模型数量最大值, 不用修改
	val_interval	进行验证的间隔
	resume	从已保存的模型恢复训练
test.py	path	存放测试图像的文件夹路径, 不用修改
	epoch	载入模型的轮数, 默认 10
	save	是否保存测试结果
visualize.py	path	存放可视化图像的文件夹路径, 不用修改
	epoch	载入模型的轮数, 默认 10
	type	可视化类型, 可以设为 filter, feature 或 tsne
	layer_idx	可视化卷积层层数, type 设为 filter 或 feature 时使用, 可以为 0、1、2、3、4
	image_idx	可视化图片索引值, type 设为 feature 时使用
	save_dir	可视化结果存放文件夹路径, 不用修改
cal_entropy.py	mode	熵计算类型, 可以设为 dataset 或 single
	train_file_path	存放训练图像的文件夹路径, 不用修改
	test_file_path	存放测试图像的文件夹路径, 不用修改
	im_path	用于熵计算的图片路径