
華中科技大學

課程報告

課程名稱： 機器學習導論

專業班級： CS2303

學 號： U202315494

姓 名： 王杰瑞

計算機科學與技術學院

一、任务目的与要求

使用 Fashion-MNIST 数据集，构建图像分类模型，将不同类别的服饰（如 T 恤、外套、鞋子等）进行识别。此数据集是经典 MNIST 数据集的替代品，图像为 28 × 28 灰度图。

二、任务方案

采用多种 CNN 模型对数据进行拟合，使用最简单的 ERM 方案，并对参数和数据处理做控制变量实验，来探究其对拟合效果的影响。

三、数据分析与处理

下载数据集之后，解压缩之后形成 ImageFolder 所需的目录结构，在运行时，使用 PIL.Image 加载并转化为 Tensor。一半的实验使用数据加强后的 Tensor，经过了随机反转，正则化的处理。

数据集链接：<https://github.com/zalandoresearch/fashion-mnist>

四、模型构建

首先构建最基本的 CNN，然后引用 torchvision.models 里面的 ResNet18，ResNet34，ResNet50 和 VGG11_BN。MLP 的构建模式如图 1。



图 1 CNN 模型结构图

五、实验

实验的参数包括：model 模型，batch_size 批次大小，lr 学习率，pretrained 是否采用预训练的模型（CNN 无效），augment 是否采用数据增强。

其中 model 设置为[CNN，ResNet18，ResNet34，ResNet50，VGG11_BN]，lr 设置为 [0.001,0.0001]，batch_size 设置为[64,256]，pretrained 设置为[False,True]，augment 设置为 [False,True]。

一共进行 80 组实验，每组实验训练 50 个 epoch，采用 ERM 方法，设置固定的随机数种子 0。

详细的实验结果见文件 analysis.csv，在这里给出每种模型在标准参数下的结果，如下表所示。

模型	准确率（%）
CNN	89.64
ResNet18	91.61
ResNet34	91.8
ResNet50	91.71
VGG11_BN	92.27

表 1 batch_size=64, lr=0.001, pretrained=False, augment=False 下的结果

分析实验结果，可以得到以下结论。

1. 学习率会影响拟合的速度，设置过小会导致学习效果降低

学习率	准确率（%）
0.001	89.64
0.0001	89.08

表 2 batch_size=64, model=CNN pretrained=False, augment=False 下的结果

2. 使用数据增强时，未必会导致结果更好

数据增强	准确率（%）
False	92.27
True	41.74

表 3 batch_size=64, lr=0.001, model=VGG11_BN, pretrained=False 下的结果

由于实验数据比较简单，采用随机翻转和正则化的方式处理会导致每次训练的样本之间差异过大，模型学习到的内容经常变化，而测试集没有很复杂，导致模型准确率大幅降低。

3. 使用预训练的模型时，学习效果更好

预训练	准确率（%）
False	91.61
True	92.1

表 4 batch_size=64, lr=0.001, model=ResNet18, augment=False 下的结果

采用预训练后的模型可以提供相比随机初始化更好的权重，同时收敛速度更快，可以增强学习的效果。

4. 复杂的模型效果未必比简单的好，存在过拟合问题

模型	准确率（%）
ResNet18	91.61
ResNet34	91.8
ResNet50	91.71

表 5 batch_size=64, lr=0.001, pretrained=False, augment=False 下的结果

从表 5 中可以得知，深度最深的 ResNet50 的效果相比 ResNet34 反而更差，说明产生了过拟合现象。由于任务比较简单，采用复杂的网络反而会降低学习效率，在实际训练的过程中，需要根据任务规模和复杂度来挑选合适的模型。

六、结论和改进

从实验的结果来看，采用 VGG11_BN 模型来完成任务有着最高的准确率 92.96%，但依旧存在可以改进的空间。比如说数据的处理方式可以探索更多的增强方式，模型的选择也可以构建更优的 CNN，采用比 ERM 更高效的训练方式等等。