

Homework 9

张思源 21110850018

November 25, 2021

1 Ex1

在导入 MNIST 数据集后, 对于第一问, 采用 sklearn 自带的 MLPClassifier 进行试验. 为保证实验效果不受宽度的影响, 这里控制变量, 对每层网络均采用 50 个神经元的宽度, 并可以得到分类准确率如下图 (具体代码见 fcn.py 文件):

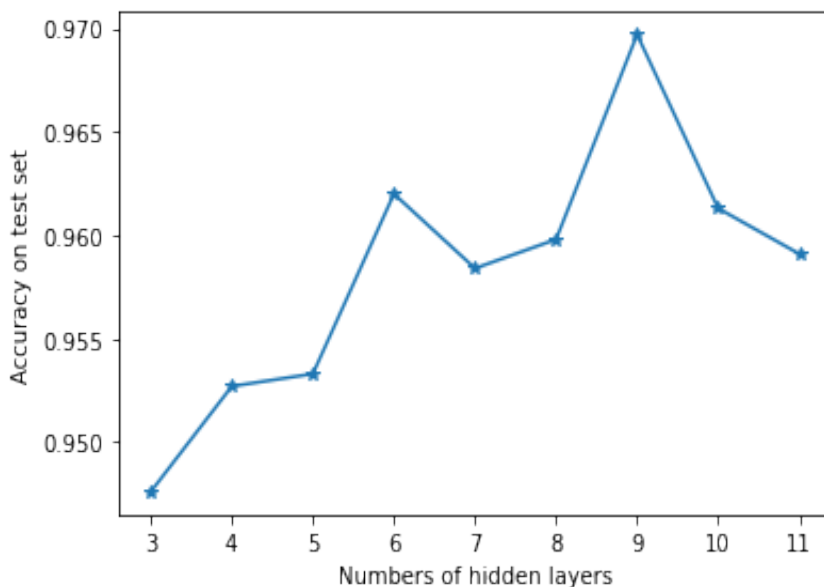


Figure 1: Accuracy varies with depth

可以看出, 在实验开始时, 测试集上的精确率随着网络深度的增加而增大, 并在网络深度达到 6 层时达到最大, 这是因为随着深度的增大, 网络的参数量增加, 网络的性能更好, 对数据的拟合效果也会更好. 但是随着深度的继续增加, 网络的精确率出现先减小后增大最后又减小的情况, 这里尝试给出解释: 网络精确率的第一次下降可能是由于随机初始化的影响, 因为此时精确率并没有下降太多且很快出现回升态势, 而网络精确率的上升是因为深度的进一步增加使得参数进一步增多, 网络的性能进一步增强, 并在层数达到 9 时达到最大; 最后再层数超过 9 后, 由于参数量过多, 网络开始出现过拟合的现象, 从而在测试集上的精确率开始降低. 从本次实验中可以得到这样一个启示, 即不可过分追求网络深度, 要结合数据集的特点, 选择合适深度的网络, 或者进行过合理的网络架构搜索, 以选取最优的网络架构.

2 Ex2

对于第二问, 这里对三种不同的网络架构, 研究宽度对其分别的影响, 为保证实验结果不受深度影响, 这里分别固定其深度, 每次改变全连接层的宽度, 以探究不同参数量下的网络精确率的变化, 并可以得到分类准确率如下图 (具体代码见 `fcn_and_cnn.py`):

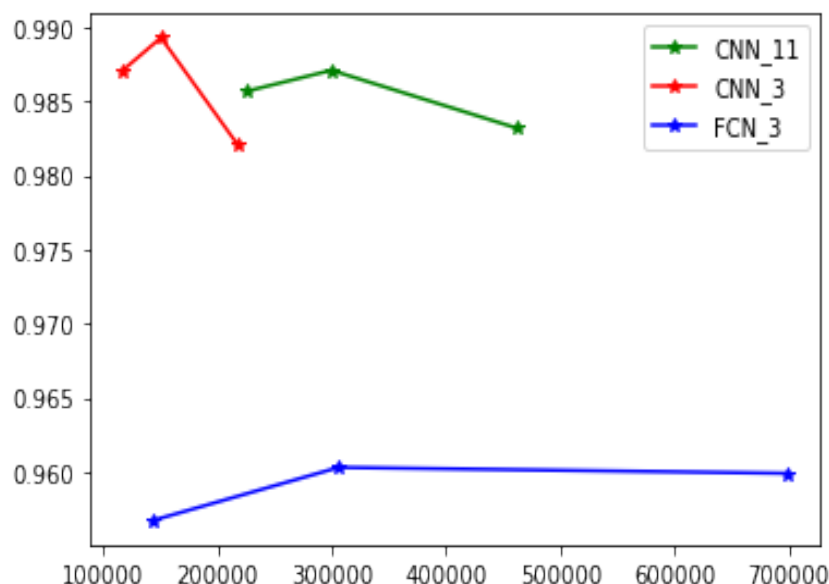


Figure 2: Accuracy varies with width

可以看出, 对于三种网络, 均呈现出随着参数量增加, 网络在测试集上的精确率先增后降的态势, 这与 Ex1 中的分析类似, 是因为网络随着参数量的增加性能先是增强, 而后出现过拟合导致的. 再比较三种网络之间的差异, 可以看出 CNN 的表现明显优于 FCN, 这是因为 CNN 的卷积操作, 保留了图像的局部结构信息, 更适合处理图像信息, 而 FCN 直接对图像的矩阵进行向量化操作 (即“拉平”), 会导致图像局部信息的丢失. 而同样是 CNN, 3 层卷积网络的性能甚至要优于 11 层网络, 这一点较为反常, 这里尝试给出解释: 因为 MNIST 数据集中的图片像素为 28×28 , 尽管在程序中采取了类似 LeNET 中的 `Resize` 操作将其重塑为 32×32 的图像, 但是其尺寸仍旧较小, 随着卷积操作和池化操作的不断进行, 图像的尺寸不断缩小, 进而图像的特征不断缩小, 最终导致网络抽取捕捉的特征相对较少, 也就导致了网络的性能较 3 层卷积网络稍差. 但是尽管如此, 可以看出 11 层卷积网络的对于参数量或者说宽度的鲁棒性较 3 层卷积网络更好, 这是因为网络的深度更大, 从而对于宽度的改变相对没有那么敏感, 或者说更深的网络对于参数改变的影响更“光滑”.

References

- [1] 李航. 统计学习方法 [M]. 清华大学出版社, 2012.
- [2] Goodfellow, Ian, et al. Deep Learning[M]. MIT Press, 2016.
- [3] Peter Harrington, 李锐…[等. 机器学习实战 [M]. 人民邮电出版社, 2013.