

5 分 问题 1 如果我们尝试做一个 28*28 的手写体 0-9 分类识别问题，如下数据有哪几个问题？

应该怎么改进？

label 应该为 0 - 9 共 9 类，24 不是我们的 label 应当将这一行删除

#null 空值，应该将它替换为 0 或者像素平均值

277 超出了 feature 预设，0-255，应当将它改为 255 或者将这一行删除（合理答案均可）

** 第一行均是 pixel1 不影响处理结果，是干扰项，如果同学们回答这个地方，可以不扣分。

label	pixel1	pixel1	pixel1	pixel1	pixel1
1	#null	0	0	0	0
2	0	0	0	0	0
8	0	33	0	0	0
4	0	0	255	0	277
5	0	1	0	0	0
24	0	0	0	0	0
1	0	0	0	0	51
9	0	0	0	44	0
0	0	0	0	0	0
3	0	0	0	0	0

5 分 问题 2 拿到数据后，我们经常需要通过列出数据的整体形势，比如总数，均值，label 个数，等 请大家简单描述 1-2 个函数能够反应数据情况。画图函数也可以

问题 2：

常用 `X_test.isnull().any().describe()`

`g = sns.countplot(Y_train)` 等画图函数。

同学们只需要回答出函数的意义，例如反应出均值，label 分布，数量，即可。

10 分 问题 3 混淆矩阵的四个模块是啥？

根据下列描述，算出四个模块的数值，请描述清楚每一项对应的值。

假设条件：有 200 个人，其中我们确切地知道，其中有 10 个人患有疾病 A；

这 200 个人去医院做疾病 A 的检测，得到如下结果：

医生认为有 15 个患有疾病 A 的人，其中有 8 个是真的患有疾病 A；其余 185 医生认为没有患病。就该医生的检测结果就行分析，计算混淆矩阵的 4 个值。

TP 真阳性 = 8，TN 真阴性 = 183，FP 假阳性 = 7，FN 假阴性 = 2

10 分 问题 4 两条 roc 曲线指示了两个模型的优劣，请问应该用哪一个单一数学值，来衡量两个 roc 曲线的优劣？

AUC area under the curve，曲线下面积，越大越好

10 分 问题 5 如果我们有一个线性回归问题，给定若干点，拟合一条一元一次线性函数

$$Y = WX + b$$

用文字或伪代码描述梯度下降法寻找 W 和 b 的过程

伪代码：

学习的函数/模型 $y = WX + b$

loss = 任意损失函数，合理即可

$y_{\text{predicted}} = WX_{\text{train}} + b$

$Y_{\text{true}} - y_{\text{predicted}} = Y_{\text{true}} - WX_{\text{train}} - b = \text{Loss}$

for i in range(epoch_num):

$Y_{\text{true}} - y_{\text{predicted}} = Y_{\text{true}} - WX_{\text{train}} - b = \text{Loss}$

$W_{\text{new}} = W - \alpha * d\text{Loss}/dW$

$b_{\text{new}} = b - \alpha * d\text{Loss}/db$

$Y_{\text{true}} - y_{\text{predicted}} = Y_{\text{true}} - W_{\text{new}} * X_{\text{train}} - b_{\text{new}} = \text{Loss}$

实际作业练习代码：

def loss_fn(x, y):

$y_{\text{pred}} = \text{random_line}(x)$

 return tf.reduce_mean(tf.square($y_{\text{pred}} - y$))

for epoch in range(EPOCHS): # 迭代次数

 with tf.GradientTape() as tape: # 追踪梯度

 loss = loss_fn(X, y) # 计算损失

 dW, db = tape.gradient(loss, [W, b]) # 计算梯度

 W.assign_sub(LEARNING_RATE * dW) # 更新梯度 通过从 "ref" 中减去 "value" 来更新 "ref".

 b.assign_sub(LEARNING_RATE * db)

10 分 问题 6 在课上我们讲述了 CNN 卷积神经网络中有 3 个关键的层，分别是什么？

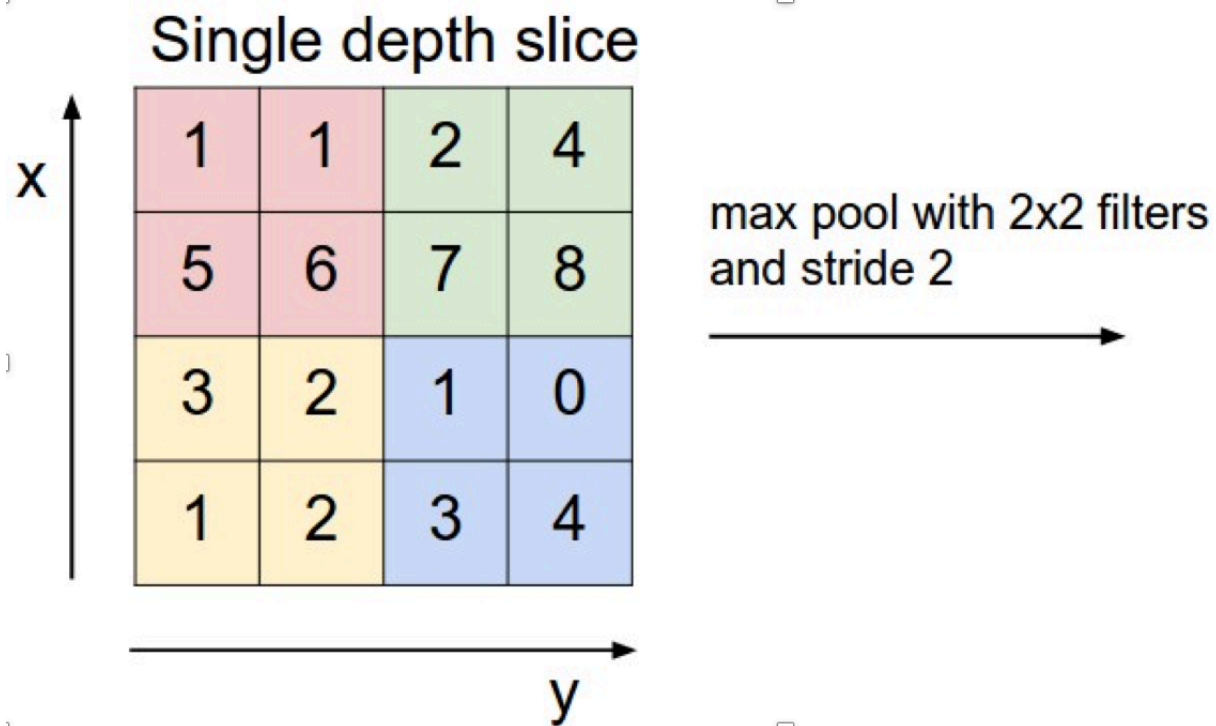
卷积层，池化层，全联接层

5 分 问题 7 输入的图片大小为 28×28 ，卷积核的大小是 3×3 ，卷积核每次扫描的步长为 1，不加 zero-padding，请问输出的 feature map 大小是多少？写出计算过程。

$$(28 - 3 + 0 * 2) / 1 + 1$$

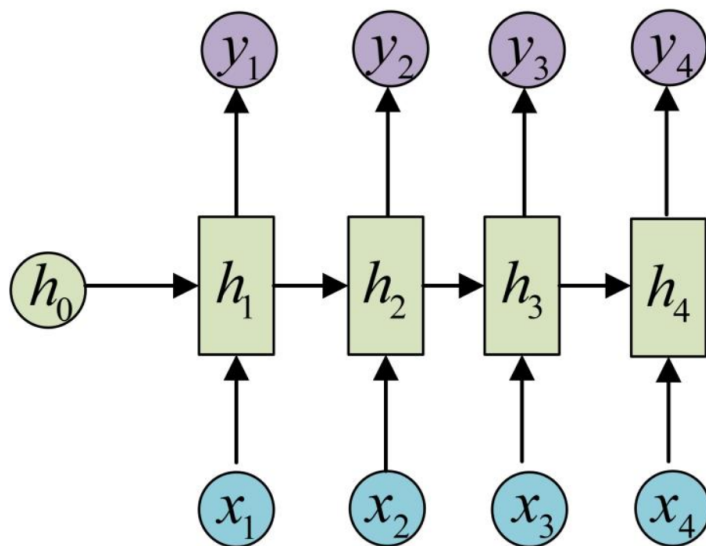
计算过程需要算分，只得出结果 26×26 可以酌情给分

5 分 问题 8 下图 feature map 经过 2×2 的 filter 和步长 2 的 maxpooling 结果是什么？



6 8
3 4

10 分 问题 9 下图是 RNN 的最常见结构，请写出 t 时刻神经元的状态函数 h_t



$$h^{(t)} = \phi(Ux^{(t)} + Wh^{(t-1)} + b)$$

答案

注意下标 t 一定要有，即 $t/t-1$

10 分 问题 10 LSTM 结构中，有几个门？为什么一般 RNN 要用 LSTM 结构？

遗忘门 Forget gate

输入门 Input gate

*注 在输入门处有 2 层 sigmoid, tanh 如果同学们回答出 2 个函数，并且认为这是 2 个门，也可。

输出门 Output gate

减少靠前时间戳下信息的流失。使得靠前的信息也可以传播到较后时间戳。

10 分 在作业中 我们使用的设置为如下：

```
batch_size = 64
num_classes = 10
epochs = 20
```

2 分 请问如果样本数量为 1280，请问整个模型被训练了几次（迭代次数 iteration）

20

4 分 请简述 epoch 是什么，用公式也可以

所有样本被训练一次为 epoch

▫ $\text{Sample size} = \text{Batch} * \text{iteration} = 1 \text{ epoch} = 1 \text{ training}$

4 分，请简述 batch_size 的作用是什么，为什么要这么做？

每次从所有训练数据中取一个子集（mini-batch）用于计算梯度

可以从过拟合，提高模型的精度等很多方面回答，合理即可

10 分

下图是作业中我们用的 CNN 模型的第一层卷积层，他的输入是 $28*28*1$ 的张量，其中有 16 个卷积核，每个核大小为 $6*6$ ，请问他的输出是几张 feature maps？有多少个参数需要训练？

```
model.add(Conv2D(16, kernel_size=(6,6),activation='relu',kernel_initializer='he_normal',input_shape=(28,28,1)))
```

输出 16 个 feature maps

$16*36 + 16 = 592$