

C3_W2 quiz

100%

1/ If you want to merge two splits 'train' and 'test' together using Splits API, how would you be able to do so?

- ☐ `tfds.load('mnist', split = pd.concat('train', 'test'))`
- ☐ `tfds.load('mnist', split = np.concat('train+test'))`
- ☒ `tfds.load('mnist', split = 'train + test')` → Passing both train and test in the string is the proper way to get both the splits.
- ☐ `tfds.load('mnist', merge = 'train+test')`

2/ The MNISTv3 dataset supports the Splits API. The train split has 70000 records in it. If you just want to create a subsplit of the first 7000 records and want to use the python slicing notation instead of Splits API, what would be the answer?

- ☐ `tfds.load('mnist:3.*.*', split='train[7000:]')`
- ☐ `tfds.load('mnist:3.*.*', subsplit='train[:7000]')`
- ☐ Read the entire train split, create a new dataset, iterate over the first 7000 of the 70000, and copy the records one-by-one to the new dataset.
- ☒ `tfds.load('mnist:3.*.*', split='train[:7000]')` → `train[:7000]` technically takes records from 0 to 6999 index value.

3/ If you want a subsplit of the first 10% of the MNISTv3 training records, what would the code look like using the Splits API?

- ☐ `tfds.load('mnist:3.*.*', split='train[:10%]')`
- ☐ `tfds.load('mnist:3.*.*', split='train[10%:]')`
- ☒ `tfds.load('mnist:3.*.*', subsplit='train[:10%]')` → `'train[:10%]'` in string format represents that we want the first 10% of the records from the train split.
- ☐ `tfds.load('mnist:3.*.*', subsplit='train[10%:]')`

4/ How many validation splits will this code generate?

`val_ds = tfds.load('mnist:3.*.*', split = ['train[{}%:{}]'.format(int(k/4),int((k+40)/4)) for k in range(0,400,40)])` → **10**

Hint: As k is incremented by 40, you get the values like (0,40), (40,80)... until the last one, (360:400). Dividing each value by 4 as you have (k/4,(k+40)/4), it will get converted to [0%:10%],[10%:20%]...[90%:100%] which is 10 splits. Note that the indices should be integers so the `int()` function was used inside the list comprehension to do that conversion.

5/ True or False : The TFRecord shards are only created for the training data.

- ☒ **False** → If your validation and test data size is bigger, they also can get sharded.

