Improving prediction

Suppose we identify a *small* sub-class of the true 8's

- which are mis-predicted
- but share a common characteristic
- that is *different* from the majority of examples that are truly 8's

For example: tilt in the "opposite" direction

What can we do to improve these mis-predictions ?

Perhaps the problem is that

- there are too few examples of the problematic sub-class
- such that the optimizer does not "focus" on getting these examples right

This is similar to the problem of *imbalanced data*

- the number of examples with Positive and Negative labels is very different

The solution to both

- this mis-labeling of a "sub-class of class" (8's titled in "opposite" direction)
- the mis-labeling of the non-dominant class in an imbalanced dataset

is the same:

- increase the number of examples similar to the mis-labeled examples
- to cause the optimizer to focus more on this sub-class

This is called *data augmentation* and will be re-visited in the module on Imbalanced Data.

In [11]:
```
print("Done")
```

```
Done
```