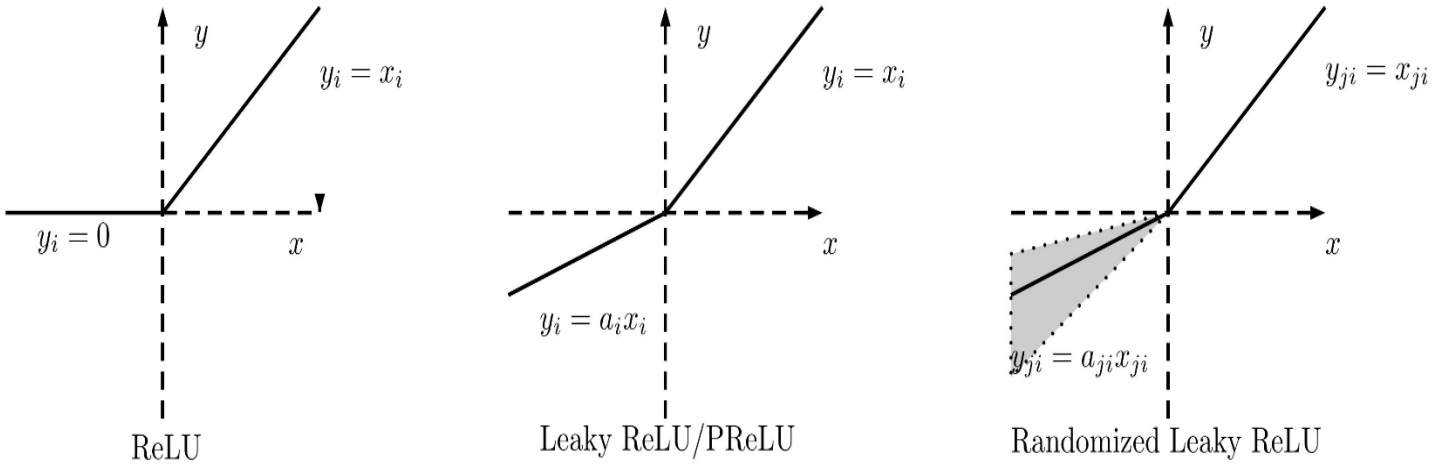
Exercise 5:

There is not much difference in appearance, but the LRELU model shows its effects in train with higher epochs, because like RELU, it does not increase the data density at 0 and tends to zero with a very small and reasonable slope, it can perform better in speed train because it has more balance and it can be the case that in more and higher repetitions, some nodes (neurons) can be turned into dead neurons so that a more appropriate performance of the model is formed.



Two advantages that LeRu has over its simpler model:

First, dying solves Relu

Second, it is faster in training.

having the “mean activation” be close to 0 makes training faster.

In fact, gradient vanishing means that the weights are not updated properly and it is so small that it cannot be trained and the model does not learn anything. Because LReLu does not set all of them to zero for smaller values and assigns them to a small value, it does not cause this problem like ReLu.