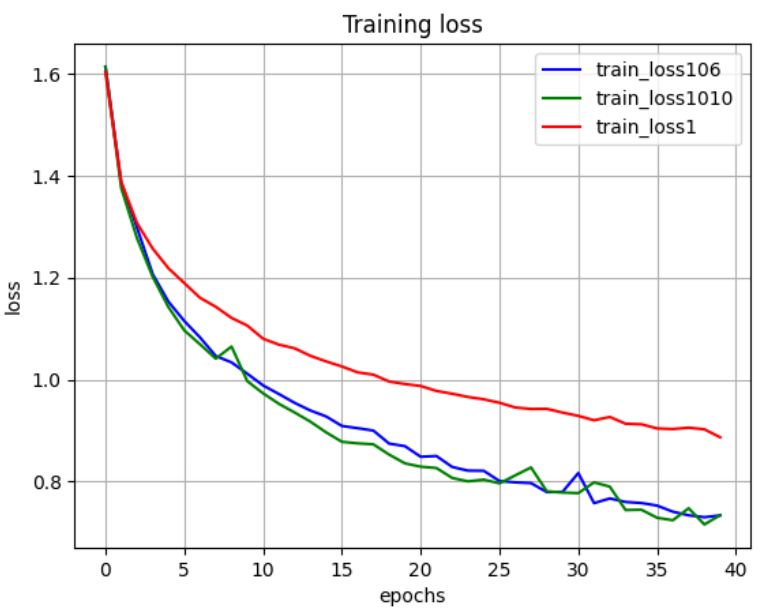
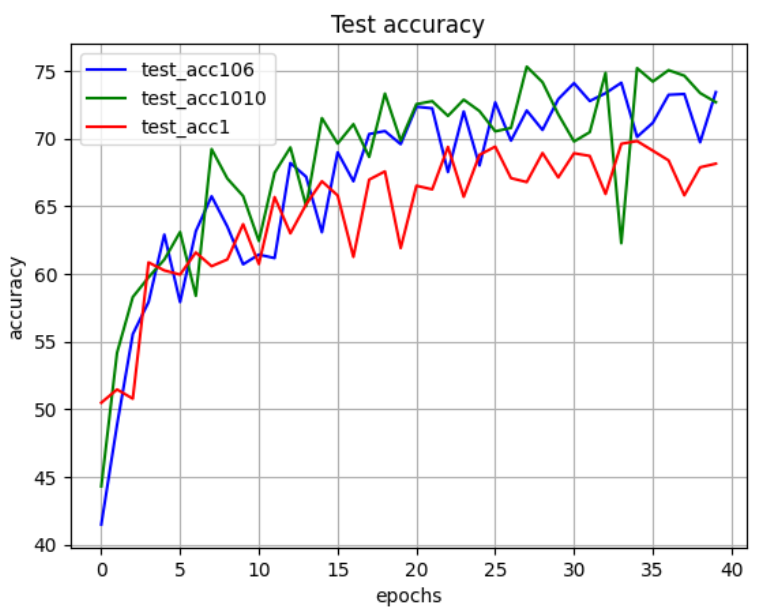
SUMMARY of the main results obtained

3 different nets are proposed in order to compare the results:

* simple\_cnn: one single convolutional nn
* not\_sparse\_MoE: mixture of 10 experts where every batch element is processed by every expert, not using the concept of sparsity (n=10, k=10)
* sparse\_MoE: mixture of 10 experts where only 6 of them are used to process each batch element (n=10, k=6)

Next are shown two plots, one for the training loss and one for the test accuracy, in order to compare their evolution during the epochs (red for simple\_cnn, green for the dense MoE, blue for the sparse one).



First of all, it is evident that the MoE models overtake (by about 5 points of percentage) the simple-cnn, as it was easily predictable.

What is surprising, instead, is the fact that the dense and the sparse model produce very similar results, but with a significant difference in the computational costs. In fact, for example, we have observed that the time needed for an epoch proportionally increases with the number of experts used for each batch element (about 180 seconds for the sparse model and 300 seconds for the dense one, using cpu).

So it is quite clear that, with equal values of loss and accuracy, it is preferable a model that requires less computational resources.