ML4Science

Week 3 meeting

R. Brioschi, G. D'Angeli, F. Di Gennaro

Tutor: Dr. Michele Bianco

EPFL, Laboratory of Astrophysics (LASTRO)

05/12/2022

EPFL

Outline

1. Improvements of NN of last time
   - Sigmoid removed
   - Different way to standardize data
2. New networks
   - NN with data of last time
   - NN with new data and masking to remove zeros
   - NN with new data without masking

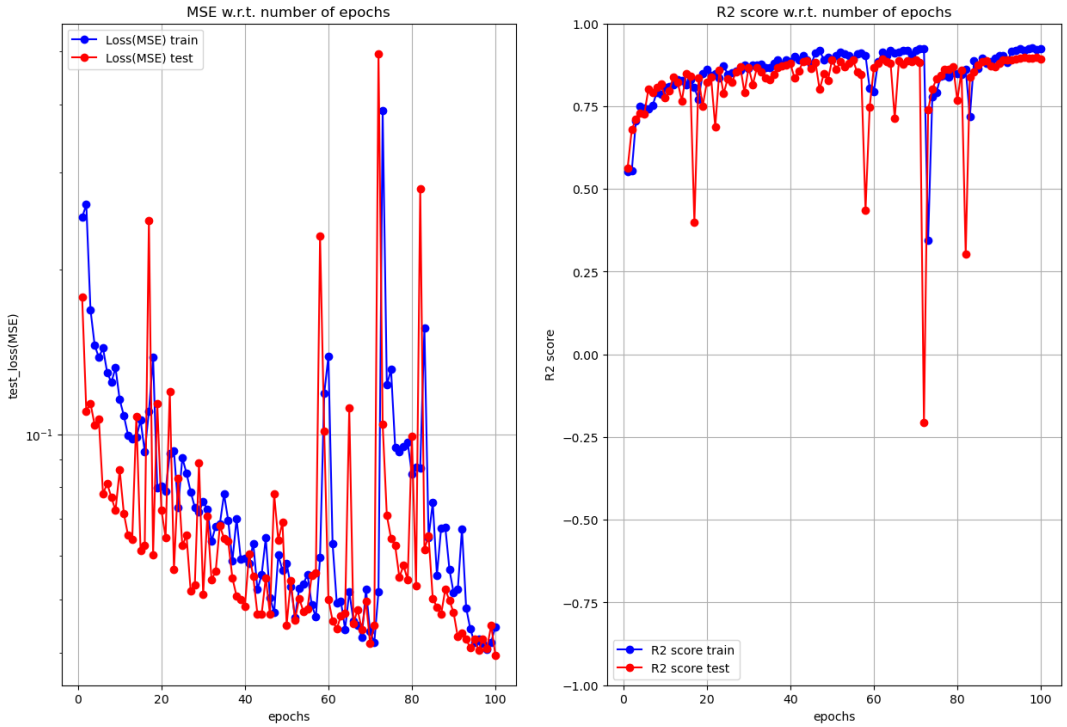
Architecture

As discussed in the email of this week, we decided to change a little bit the architecture of our network. After some trials, we understand that the following architecture works really well.

```
self.l1 = nn.Linear(num-feature,64,dtype=dtype)
self.reLU1 = nn.ReLU()
self.l2 = nn.Linear(64,128,dtype=dtype)
self.reLU2 = nn.ReLU()
self.l3 = nn.Linear(128,256,dtype=dtype)
self.reLU3 = nn.ReLU()
self.l4 = nn.Linear(256,1,dtype=dtype)
```

Results 1

We firstly decided to work with the same data as last time to have a better comparison on architectures. So we firstly fixed LH_0 and we worked with all z. $R^2_{test} \approx 0.85$.
X: Normal standardization, Y: Division for 10^{10}



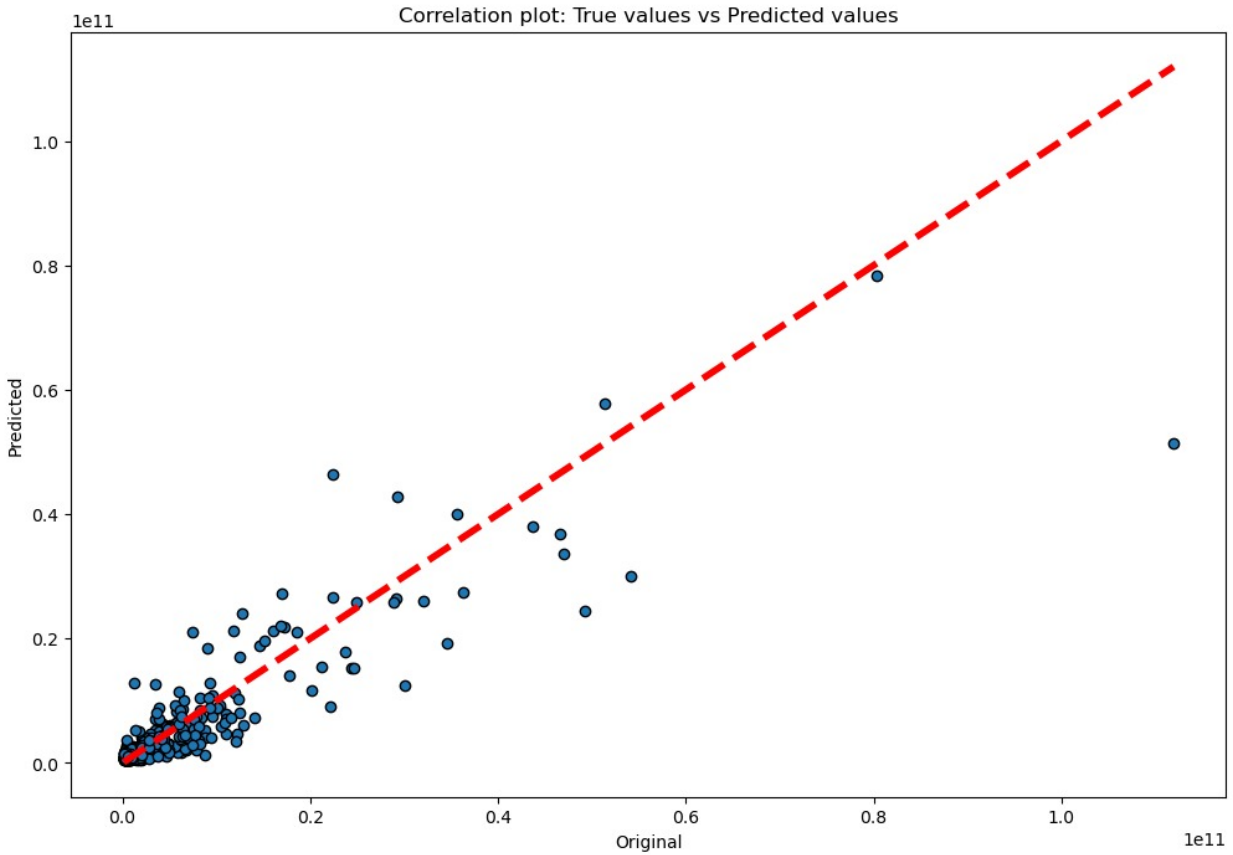
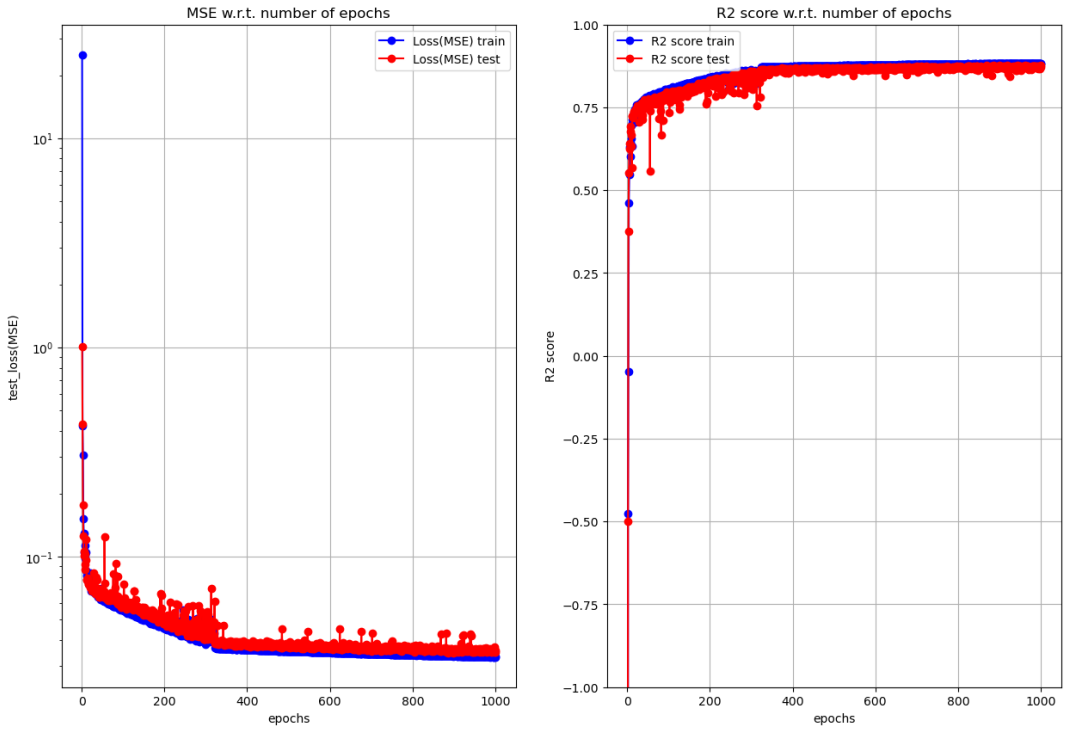


Figure 2: Correlation plot old data.

Results 2

Then we moved on new data with masking in order to remove all the zeros (still, fixed LH_0 and we worked with all z). $R^2_{test} \approx 0.85$.
X: Normal standardization, Y: $log_{10}(Y)$.



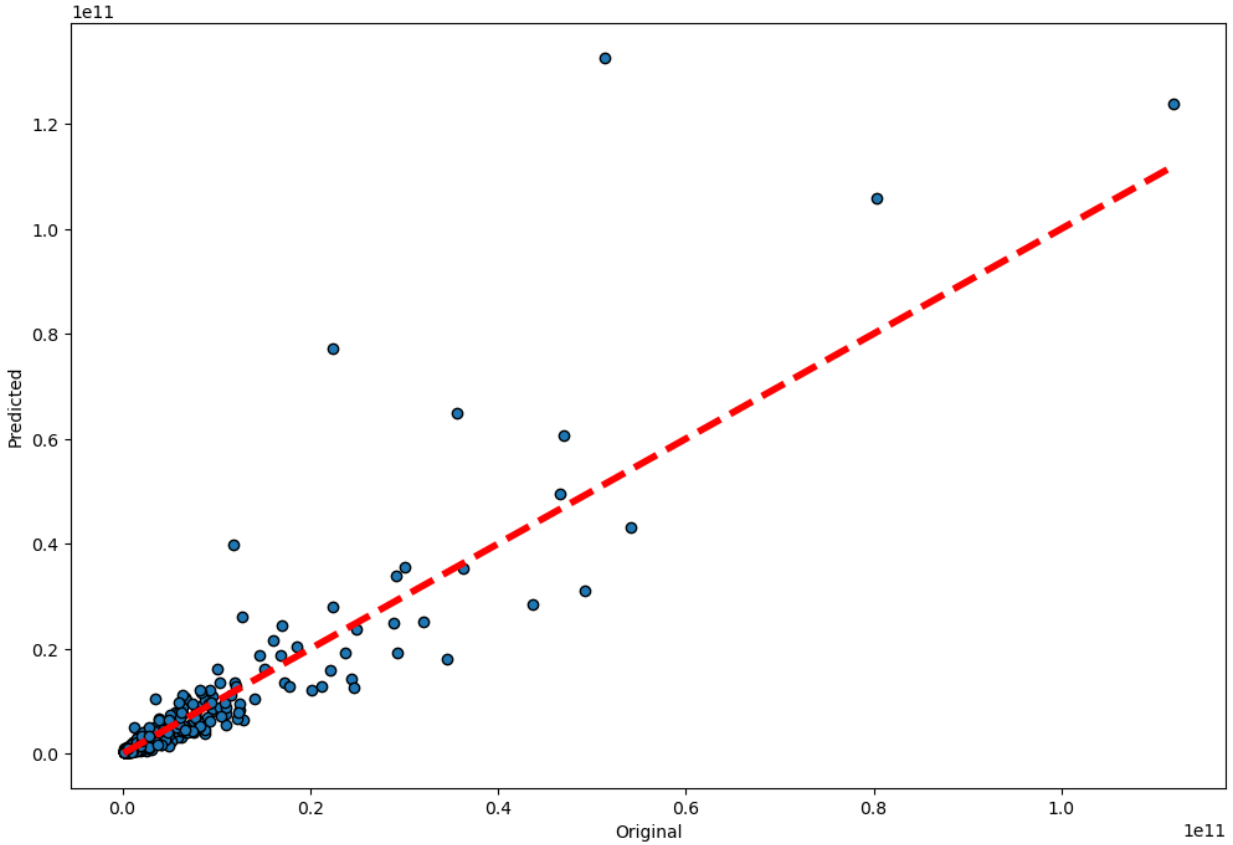


Figure 4: Correlation plot, new data and masking.

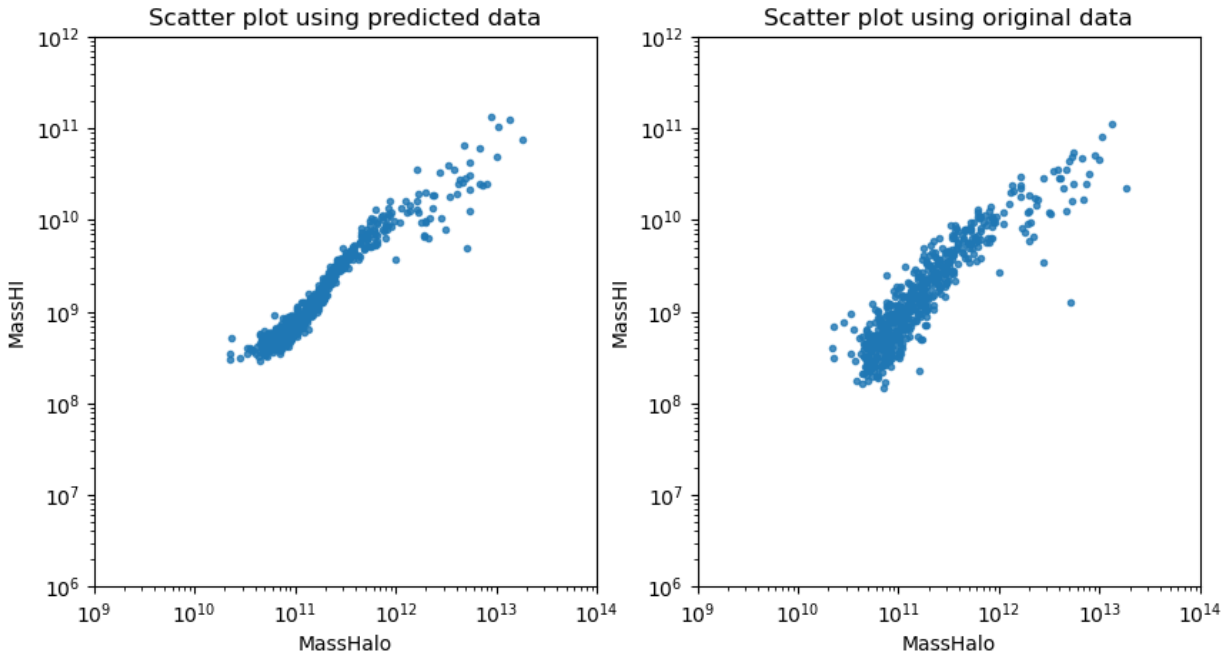
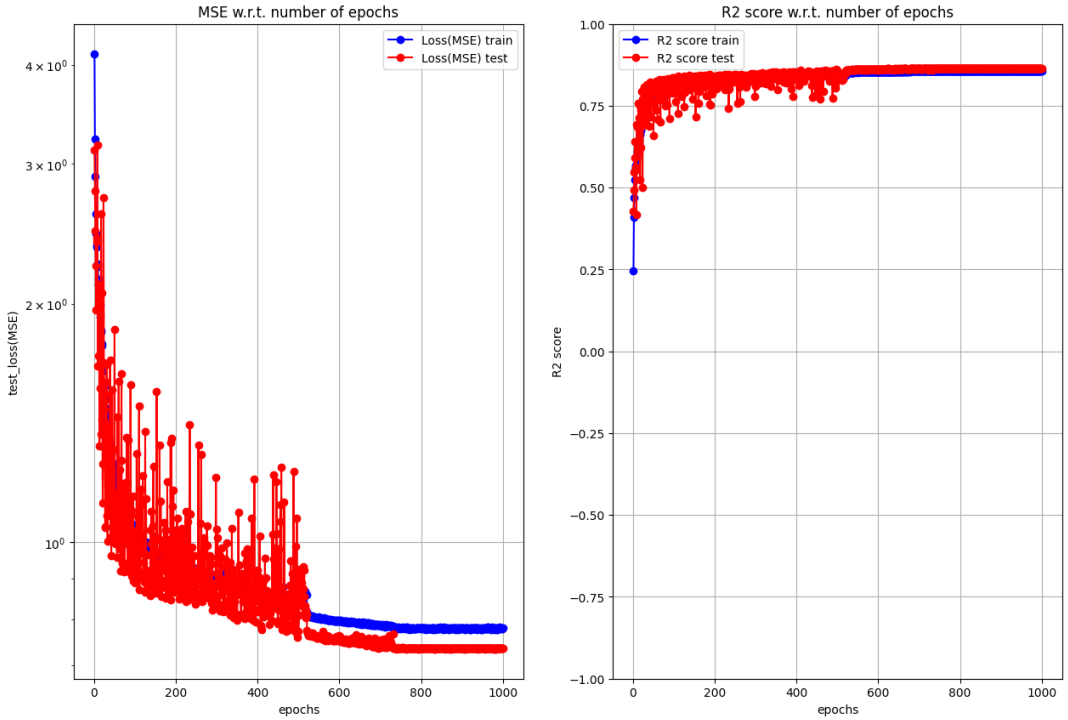


Figure 5: Cloud of points, new data and masking

Results 3

Finally, in order to generalize at the whole cloud of points, we tried to train the same model as before without masking (still, fixed LH_0 and we worked with all z). $R^2_{test} \approx 0.85$. X: Normal standardization, Y: $log_10(Y)$.



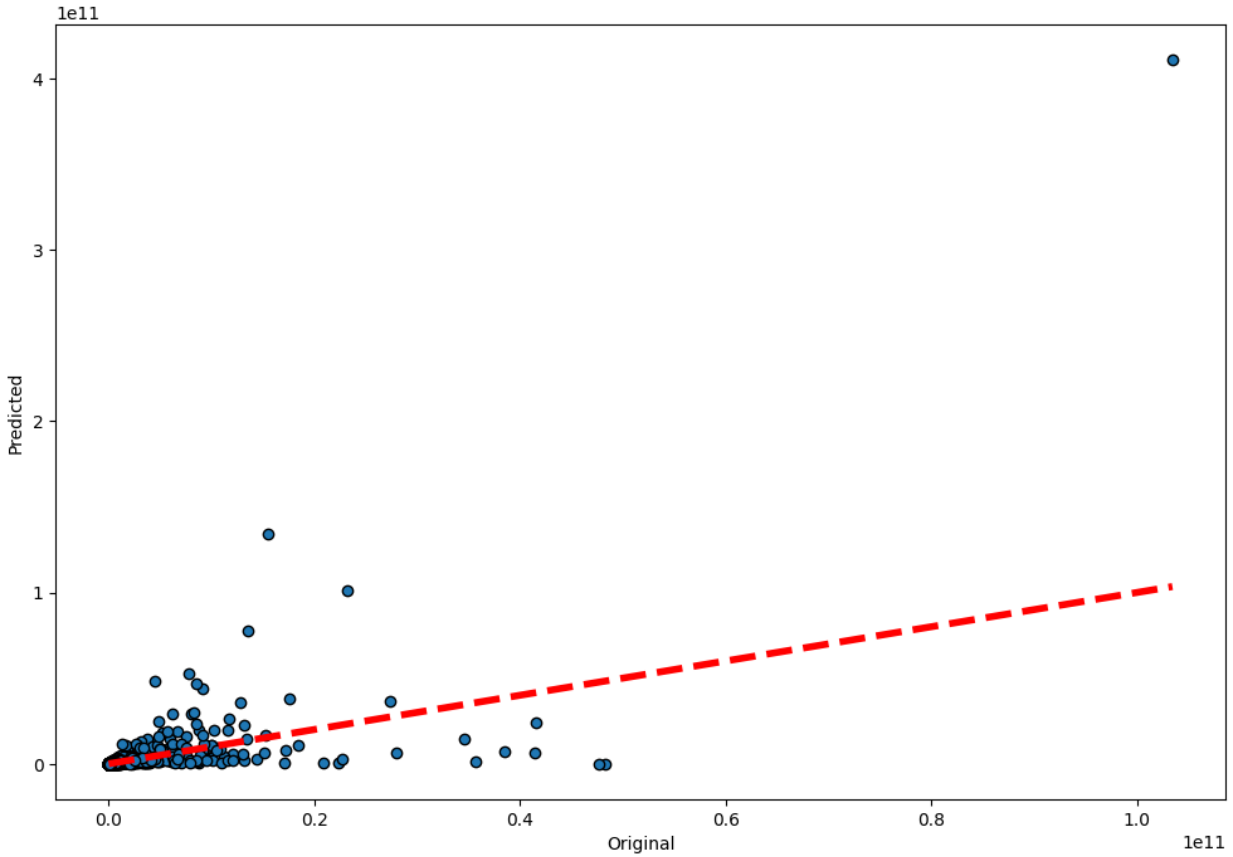


Figure 7: Correlation plot, new data and NO masking.

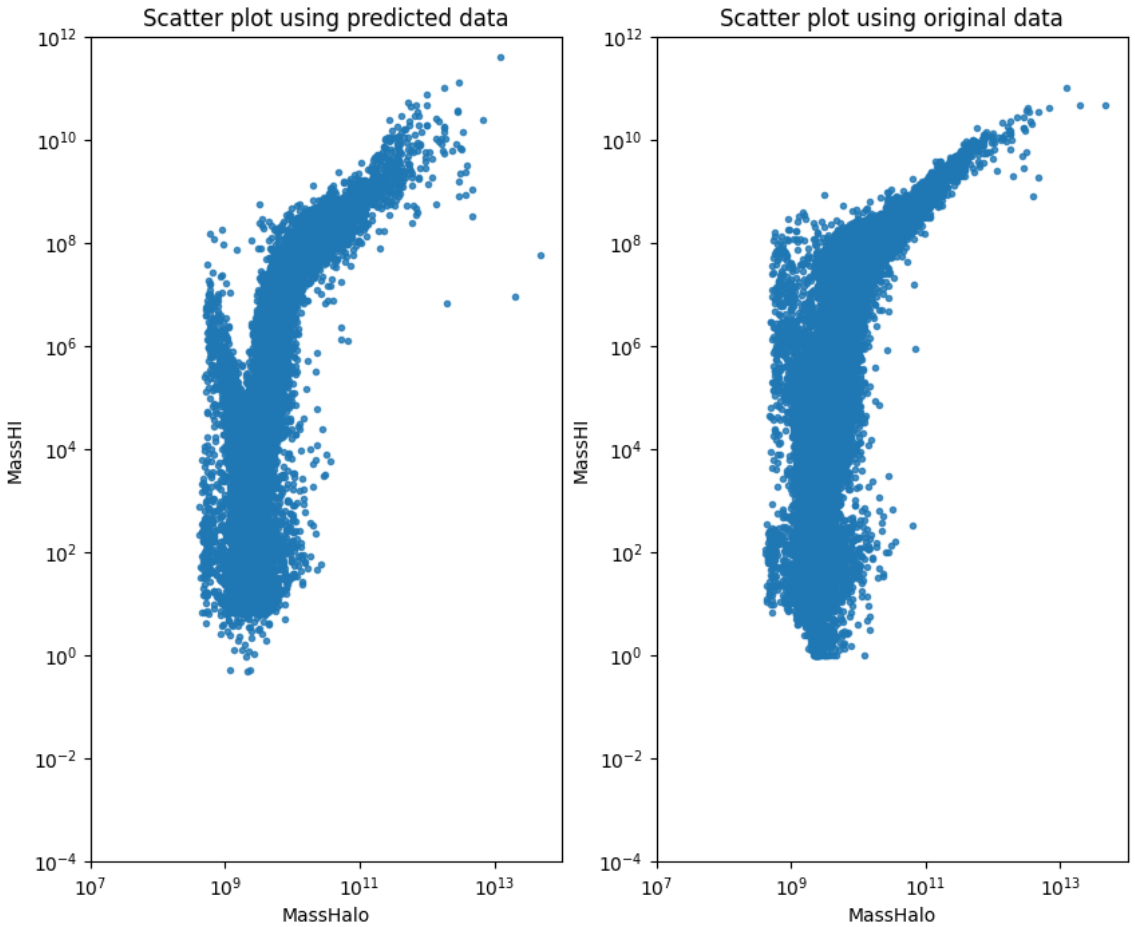


Figure 8: Cloud of points, new data and NO masking

Conclusions

From these trials we concluded that:

- Leaky Relu increases loss, slow down the train and decrease the R^2_{test}.
- Dropout increases loss, slow down the train and decrease the R^2_{test}.
- We believe that now the standardization is quite good and also the results3 tell us that our NN learn at least the unit of the output ($loss_{test} \approx 0.8$).
- The architecture *mirror* does not work well for this problem.
- This model cannot be generalized for different LH, so also the cosmological constants play an important role.