

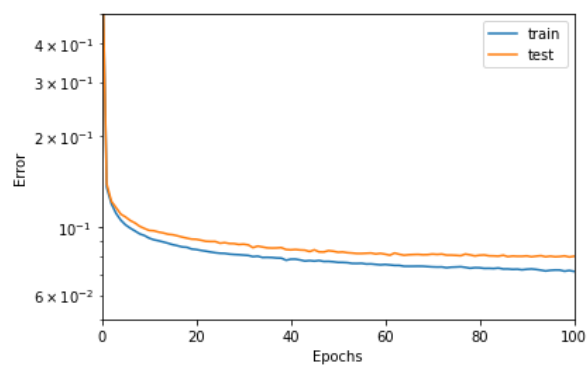
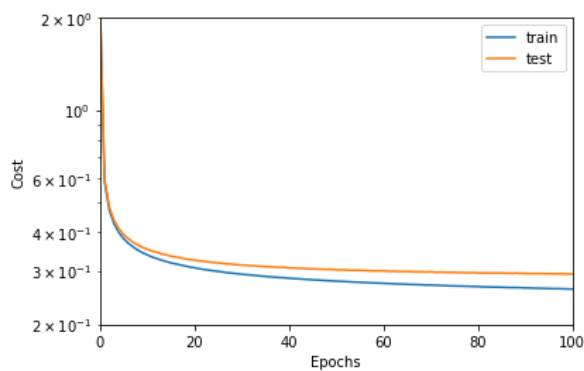
T-DeLearn - PW03

- Capocasale Romain
- Demeusy Jean
- 13.03.2021

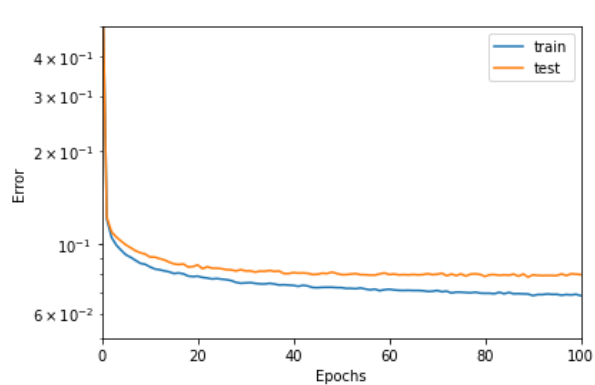
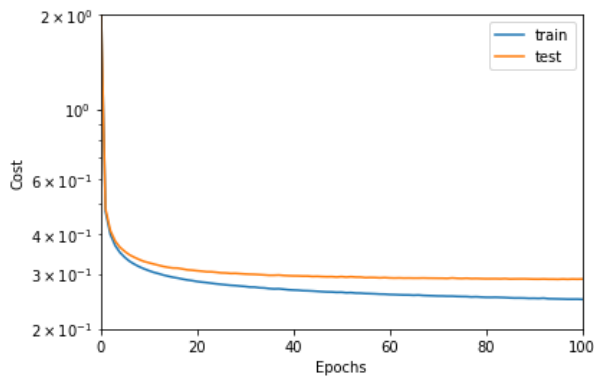
Exercise 1

Learning rate variation [batch_size=256,epochs=100]

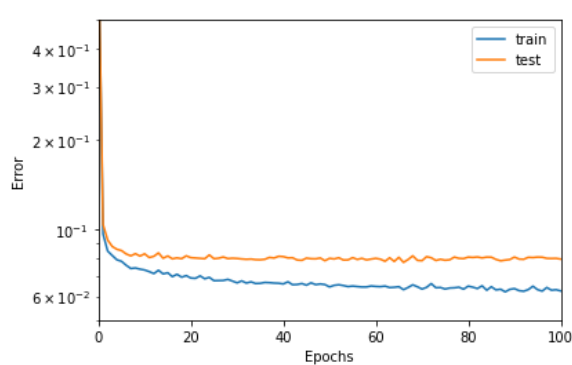
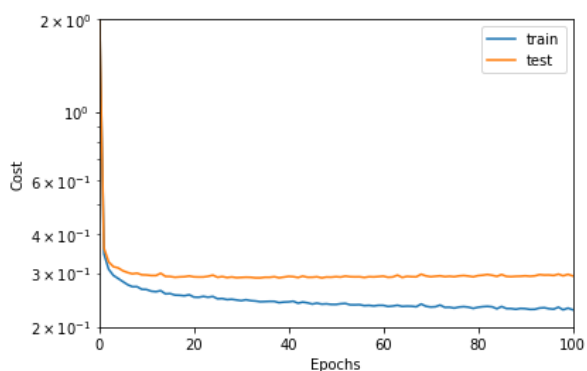
- learning rate = 0.05



- learning rate = 0.1



- learning rate = 0.5

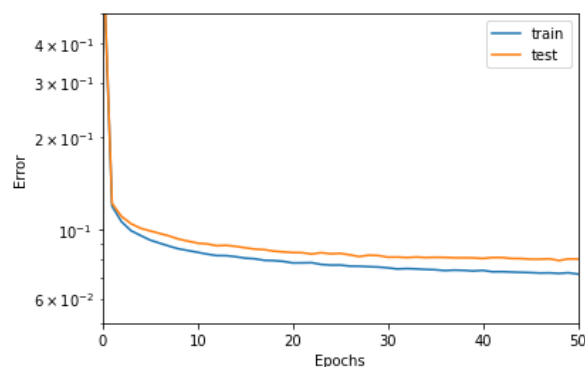
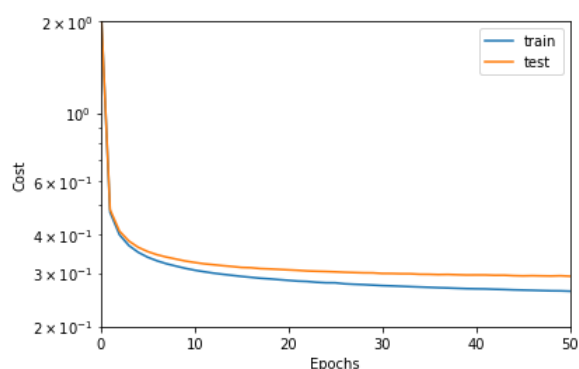


We see that the model with a learning rate of 0.5 seems to converge faster and better than the others.

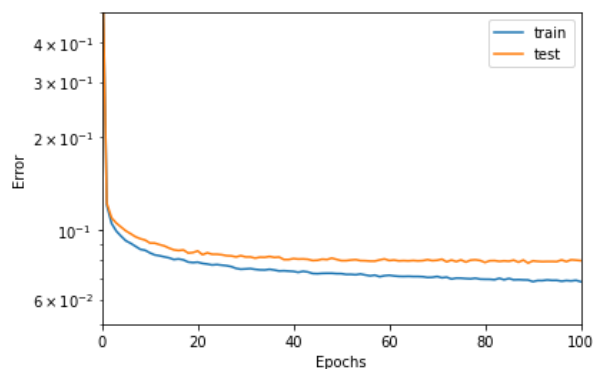
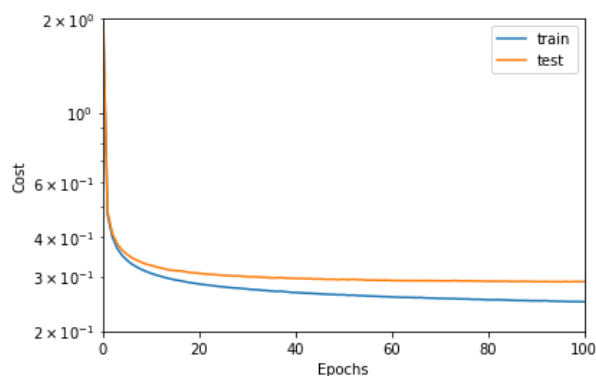
The higher the learning rate, the more the curve oscillates as opposed to the curve with a low learning rate.

Epochs variation [learning_rate=0.5, batch_size=256]

- Epochs = 50



- Epochs = 100

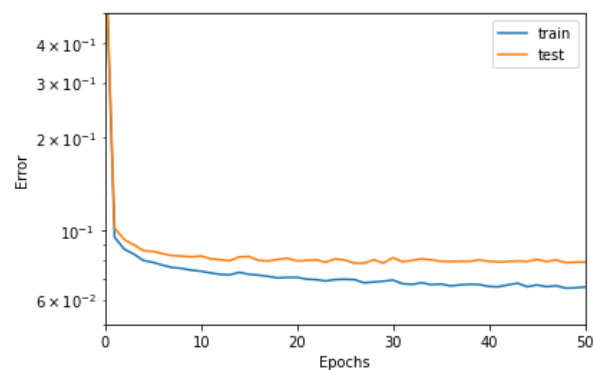
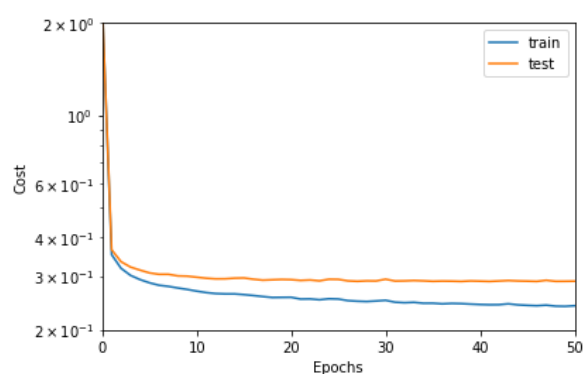


We can see that the model converges with few epochs, 50 epochs are enough to make the model converge.

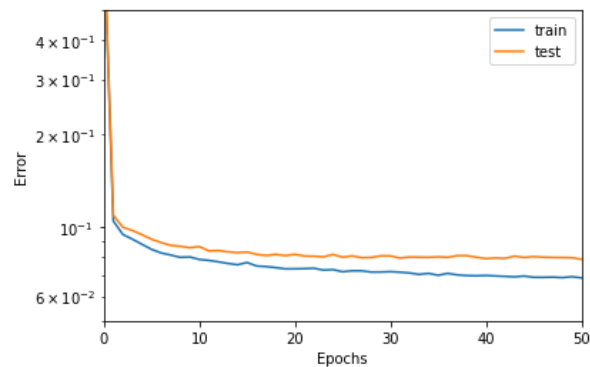
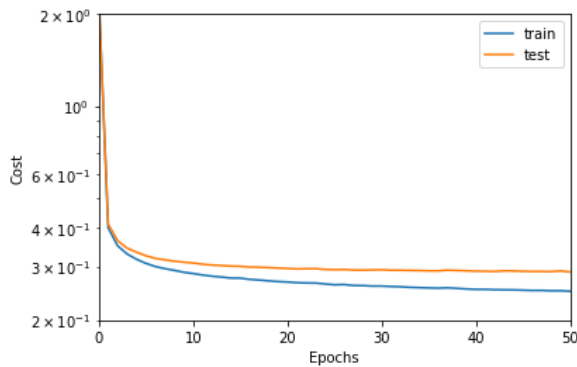
From about 20 epochs the cost and error of the test set does not go down anymore.

Batch size variation [learning_rate=0.1, epochs=50]

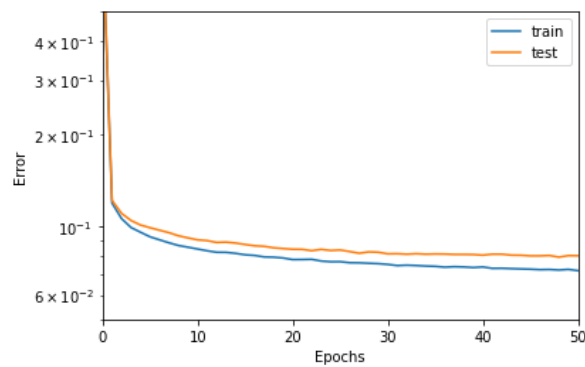
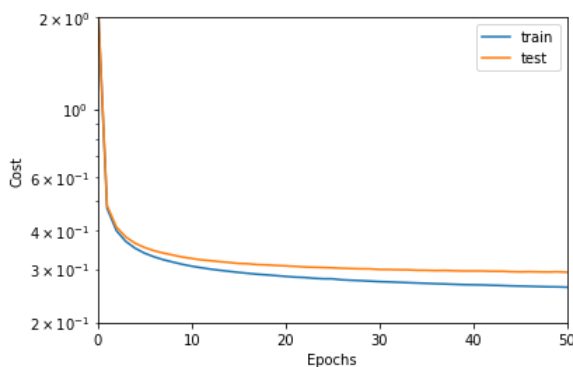
- Batch size = 64



- Batch size = 128



- Batch size = 256



Batch size does not seem to show a significant influence on the cost and error of the model

Exercise 2

a) See the appended file for gradient descent derivatives

b) See the implementation in the .ipynb file

c) When the input and/or the output data are not normalized, the model does not converge on training. The normalization of the output is mandatory because the activation function on the last layer is normalized. The normalisation of the input is necessary so that there is no problem of vanishing gradient when the range of the input is too large. With input normalization, we expect the data to have a distribution centered around 0 with a small standard deviation (1).

d) after many tries and errors, it has been determined that the best hyper-parameters are:

- epochs : 300. Under 250 epochs fitting is complicated.
- learning rate : 0.07. It has to be coordinated with the batch size.
- number of neurons : 12. Less than 10 neurons will start to give poor results.
- batch size : 2. When it increases, the cost wiggles less.

By decreasing the batch size, we had to decrease the learning rate. The number of epochs needed for the model to converge is then smaller. Also after many tries, we have determined the number of neurons can be reduced.

With the parameters proposed here, the final cost is approximately the same as with initial parameters, but the convergence is much faster.