Sheet 12 – Team MTE

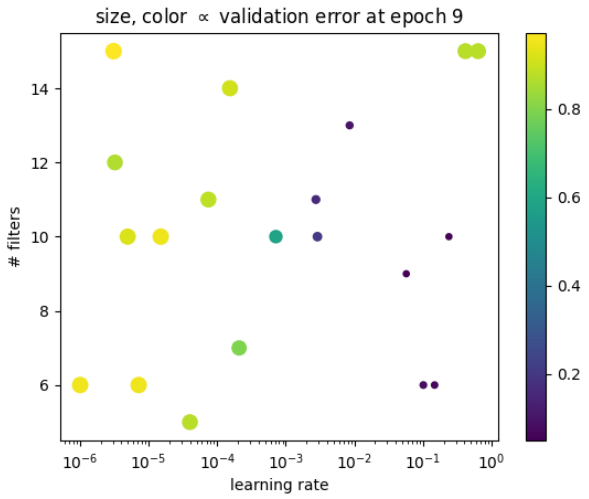
Marta Gulida - 5585808 - mg776

Erik Bode - 4505199 - kb301

Tillman Heisner - 4517815 – th273

Exercise 2.3

What pattern do you see for the scatter plot? Why might it occur?

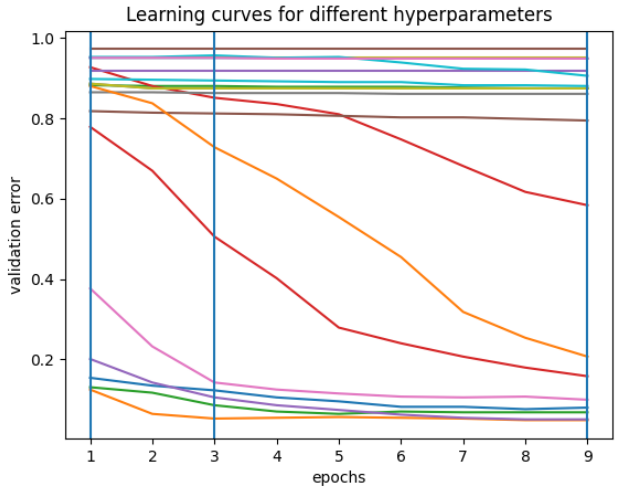


We can see that the best performance occurs when the learning rate is rather large ~ between 0.005 and 0.5. But once the learning rate gets too large, we get really bad results. This is because the U-shape curve characteristic it has about its size when tuning it.

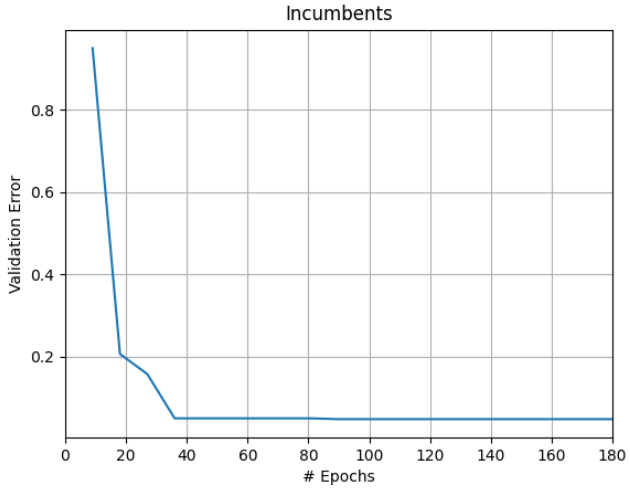
Regarding the number of filters, we can notice that they do not really matter that much. In general, we could assume that many small filters would lead to better results.

Given the error curves would you expect multi-fidelity optimization or black-box optimization to perform better?

We would assume that multi-fidelity will perform better as the error of the best hyperparameter configuration, will not decrease anymore after ~ 36 epochs

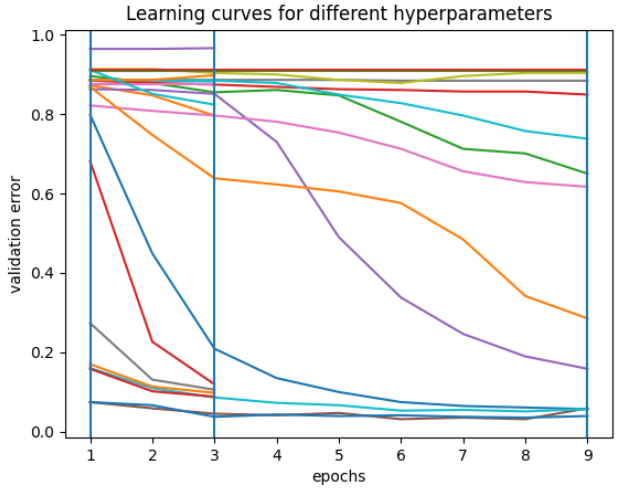


Multi fidelity may need longer to arrive at the error rates the suboptimal RS approaches reach after 2 epochs, but it will be better in the long term.

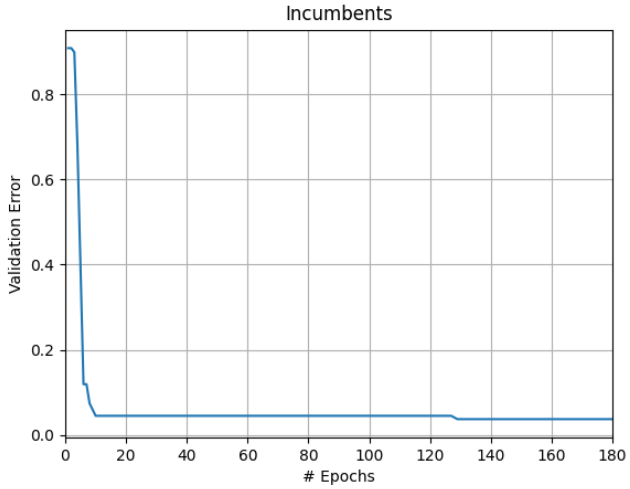


Exercise 3.3

Did HyperBand always early-stop the configurations with the worst final performance? Why?

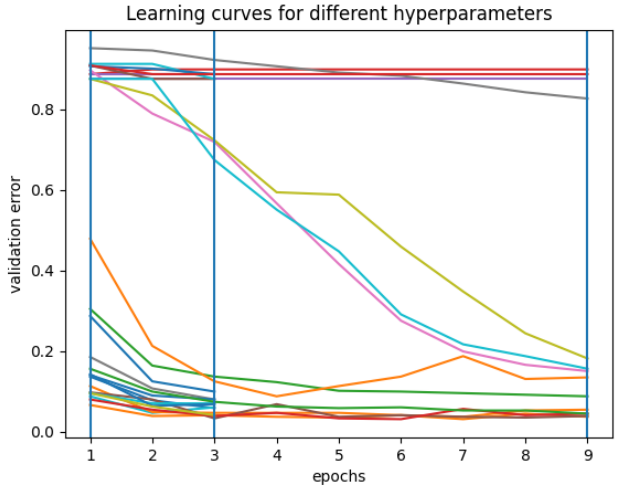


As can be seen it the picture above it did not always early-stop the worst final performance.

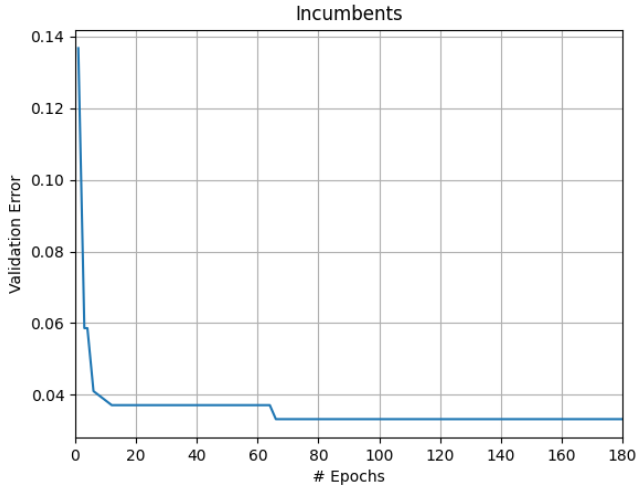


Exercise 3.5

Did PriorBand always early-stop the configurations with the worst final performance? Why?



As can be seen it the picture above it did not always early-stop the worst final performance.



Exercise 3.6

empty