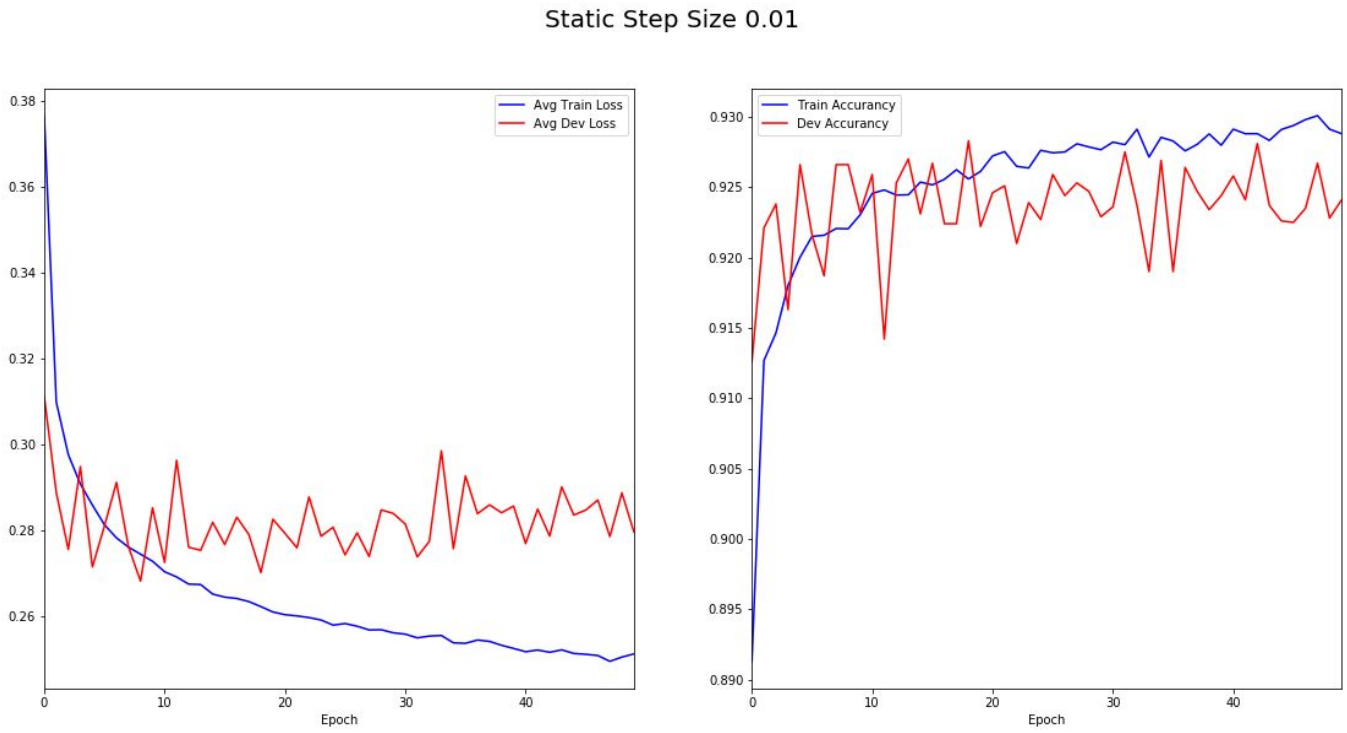
Lab Exercise 1

Edoardo CONTE <edo.conte96@gmail.com>
Pratham SOLANKI <pratham.c.solanki@gmail.com>

Model 1

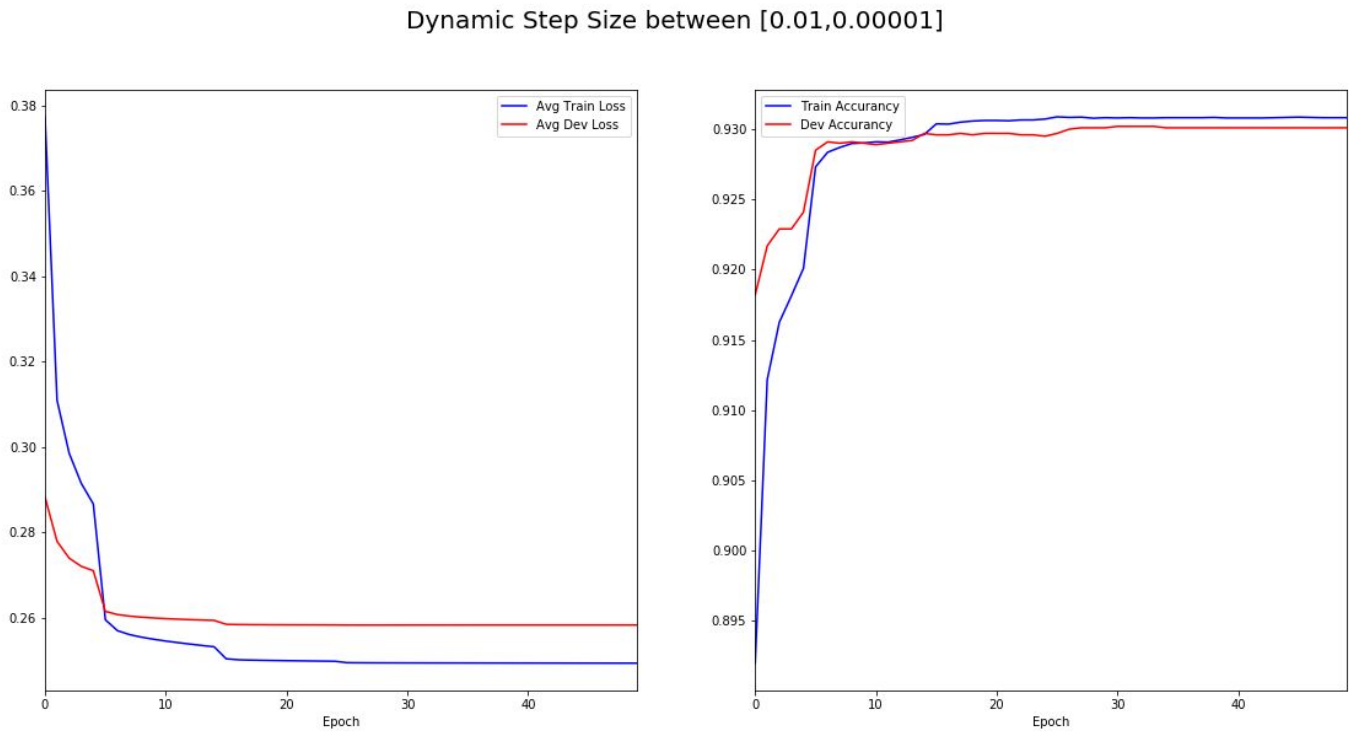
- **Static** Learning Rate: 0.01
- Number of Epochs: 50



- Observations:
    - Consistently decreasing training loss (+ increasing training accuracy), but not similar results on the dev data. Hence overfitting.

Model 2

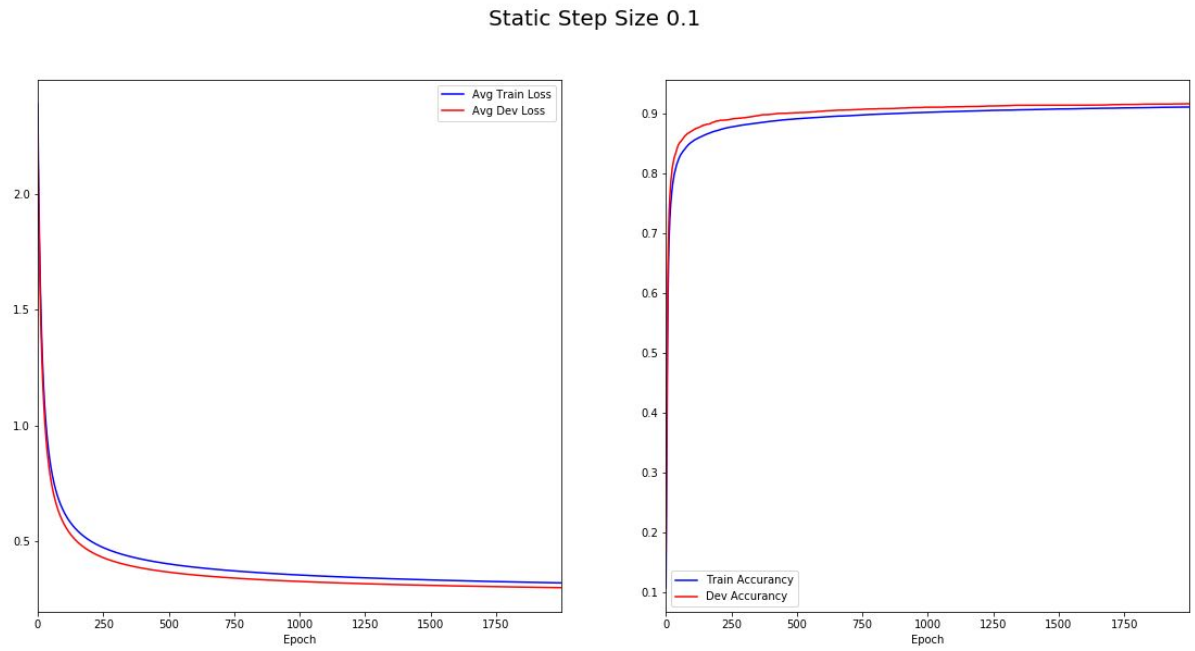
- **Dynamic** Learning Rate: {0.01, 0.001, 0.0001, 0.00001}
- Number of Epochs: 50



- Observations:
    - Decaying learning rate works perfectly and we actually converge (negligible change in loss and accuracy over several epochs).
    - Does not overfit.
    - Overall better results than the previous model with the same number of epochs.

Model 3

- **Static** Learning Rate: 0.1
- Number of Epochs: **2000**
- Parameters updated after each epoch by subtracting the average gradient



- Observations:
    - Takes too long to converge. We trained for ~5 hours and still did not converge to the global minimum.
    - However, even with a larger learning rate doesn't lead to overfitting.
    - In conclusion, the descent is very slow but is guaranteed to converge given enough number of epochs.

Comparison on Test set

Model	Accuracy	Avg. Loss
1	0.922	0.291
2	0.926	0.268
3	0.916	0.305

As we can see from the above results, the model with decaying learning rate performs the best on the test set. However note that we were not able to train the third model to convergence because of extremely slow descent.