CS 330 Autumn 2021/2022 Homework 1: Data Processing and Black-Box Meta-Learning

Kevin Lee (kelelee@stanford.edu)

Due on Wednesday October 6, 11:59 PM PST

3 Analysis

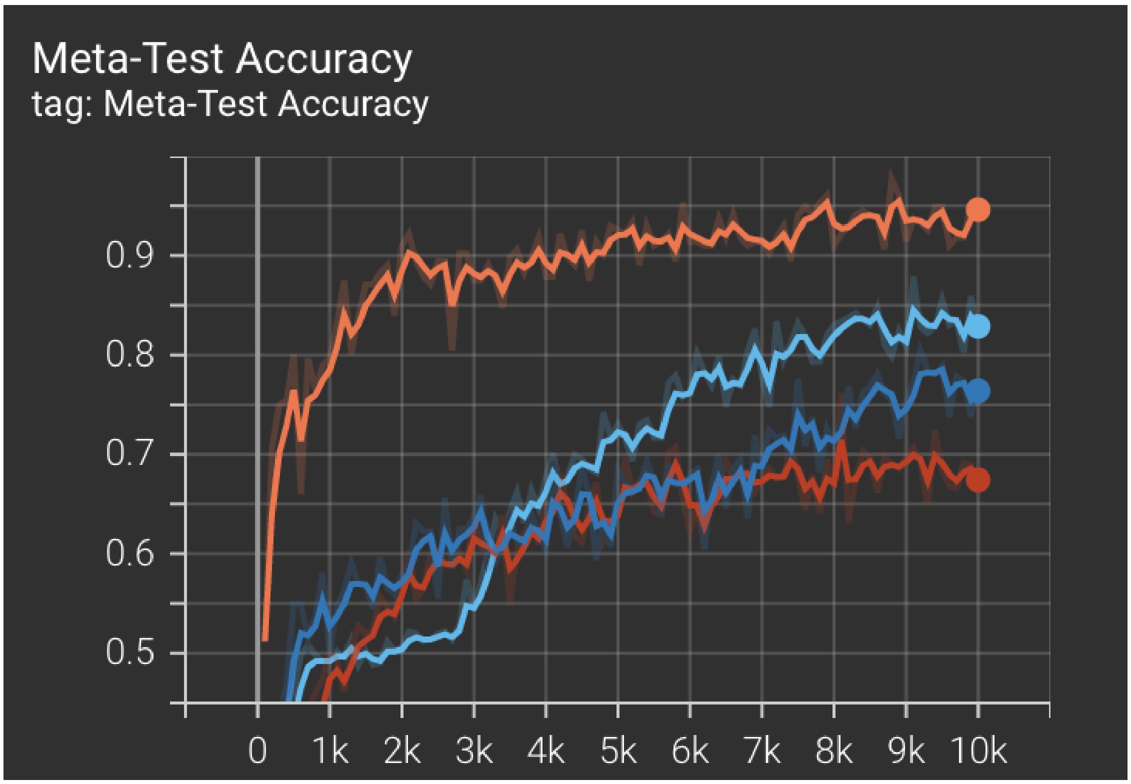

Figure 1: Meta-Test Query Set Classification Accuracy. Orange: $K = 1, N = 2$; Dark Blue: $K = 1, N = 3$; Red: $K = 1, N = 4$; Light Blue: $K = 5, N = 4$

1. How does increasing the number of classes affect learning and performance?

Answer:
With K constant, increasing N makes the problem more difficult for the model. This causes the learning and performance to become worse.

2. How does increasing the number of examples in the support set affect performance?

Answer:
With N constant, increasing K helps the model learn important features of each class better, thus increasing the model's performance.

4 Analysis

a. Experiment with one hyperparameter that affects the performance of the model, such as the type of recurrent layer, size of hidden state, learning rate, or number of layers. Submit a plot that shows how the meta-test query set classification accuracy of the model changes on 1-shot, 3-way classification as you change the parameter. Provide a brief rationale for why you chose the parameter and what you observed in the caption for the plot.

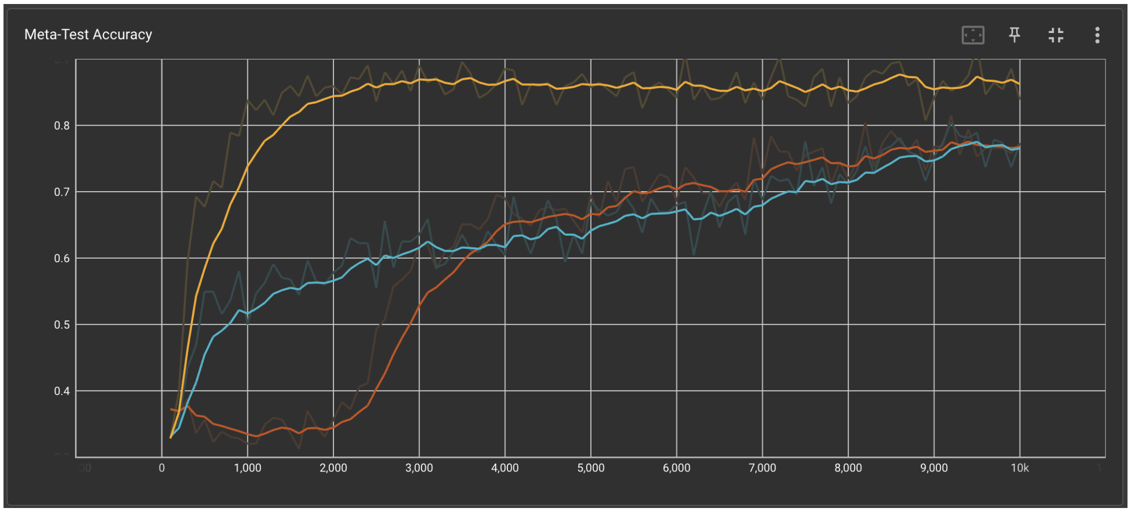

Figure 2: Meta-Test Query Set Classification Accuracy for $K = 1, N = 3$. Yellow: $\alpha = 10^{-2}$; Blue: $\alpha = 10^{-3}$; Orange: $\alpha = 10^{-4}$

Answer:
I chose to experiment with the learning rate, α, keeping $K = 1, N = 3$. With $\alpha = 1 \times 10^{-3}$, I observe that the accuracy has not plateaued at around the 10,000th step, so I tried increasing the learning rate to $\alpha = 5 \times 10^{-3}$. On the other hand, there are times when a lower learning rate is required in order to reach better local minimas, so I also tried decreasing the learning rate to $\alpha = 1 \times 10^{-4}$. It turns out that with a higher learning rate, a higher meta-test accuracy was achieved whereas with a lower learning rate, a similar accuracy was achieved.

b. **Extra Credit**: In this question we'll explore the effect of memory representation on model performance. We will focus on the $K = 1, N = 5$ case.

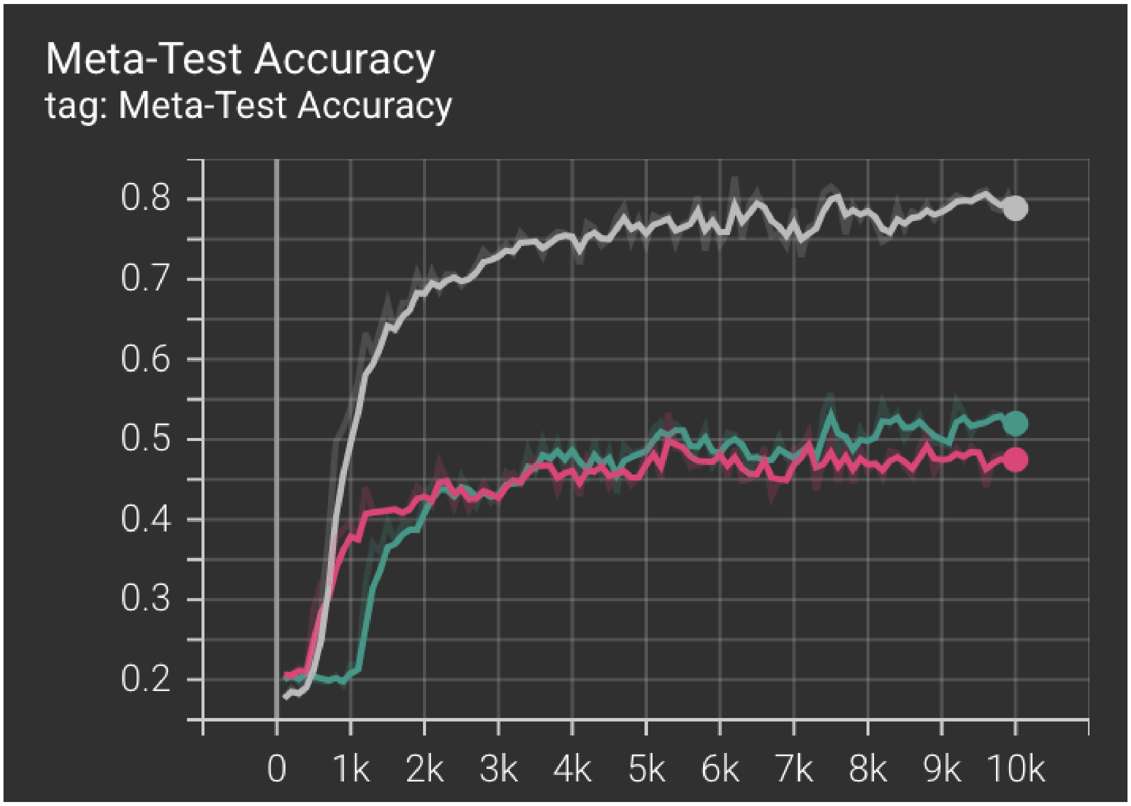

Figure 3: Meta-Test Query Set Classification Accuracy for $K = 1, N = 5$. Pink: LSTM with 128 units; Green: LSTM with 256 units; White: DNC with 128 units

b.1 In the previous experiments we used an LSTM model with 128 units. Consider a model with 256 hidden units. Does the increased memory capacity improve model performance?

Answer:
The increased memory capacity did not improve model performance significantly.

b.2 A related family of MANNs utilize a recurrent controller network, which has read-write access to an external memory buffer, allowing them to better retain and retrieve information over longer horizons. We have provided you with an implementation of a Differentiable Neural Computer model. The layer is already defined in the MANN class and you can use it as a regular recurrent layer. Can you use this model to achieve more than 80% accuracy? Compare your results with the LSTM model with 256 units. What conclusion can you draw about the different memory structures and retrieval mechanisms?

Answer:
With the DNC model, the model converges to a higher meta-test accuracy and at a faster rate. It is able to achieve more than 80% accuracy after roughly 6,000 steps. The length of the input to the models scale with K and N, and since LSTMs tend to decrease in performance when the input sequence gets too long, an architecture that performs better over long sequences such as the DNC will outperform LSTMs.