CS 330: Problem Set 1

Ji Won Park

September 30, 2020 (Due)

Colab notebook is available on

- GitHub: `https://github.com/jiwoncpark/cs330-deep-multitask-meta-learning/blob/master/hw/hw1/CS330_Homework1_Stencil.ipynb`
- Drive: `https://colab.research.google.com/drive/1NeqvyyZlaue8yKWQb0ARUg0wmqyJIrhl?usp=sharing`

See the Colab notebook for Problems 1 and 2.

3.

Fig 1 overlays the test accuracy for each of these K, N configurations. Holding N fixed, greater K (comparing $K = 5$, $N = 4$ with $K = 1$, $N = 4$) means that the network has more training support data (examples of each class) available as it builds an understanding of class, so it is more accurate at the end of training. Holding K fixed, greater N (comparing $K = 1$, $N = 2, 3, 4$) means that the classification problem is more difficult because it's more nuanced so the network is less accurate at the end of training. In particular, note that $K = 1$ is insufficient for $N = 4$, i.e. the accuracy does not improve significantly beyond the initial accuracy of 0.25 due to random initialization.

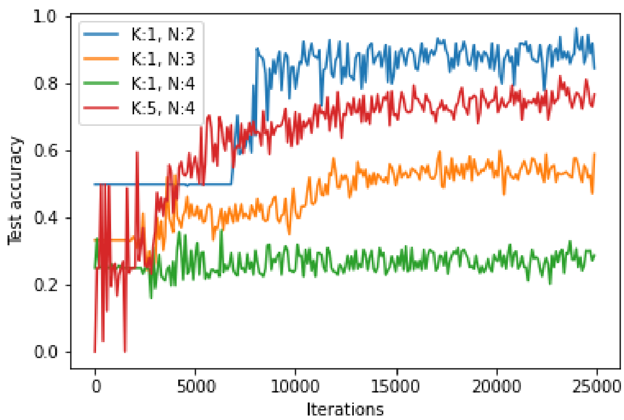

Figure 1: Test accuracy over training iterations as K, N are varied

Similarly, Fig 2 overlays the test loss over training iterations. Note that smallest N of 2 (simplest classification task) results in the smallest softmax cross-entropy loss, followed by the greatest K of 5 (more meta training support).

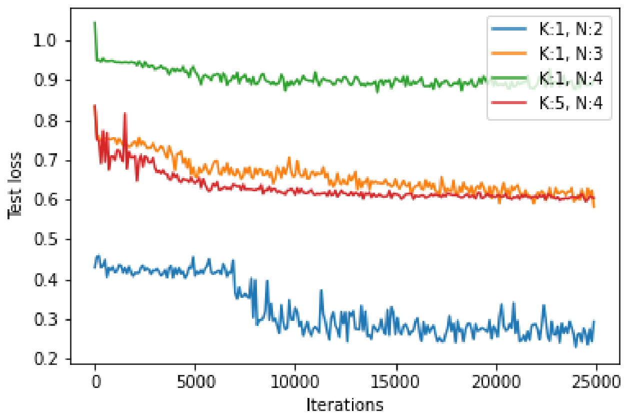

Figure 2: Test loss over training iterations as K, N are varied

4.

(a) I vary the size of the hidden state as $h \in \{64, 128, 256, 512\}$. I chose this hyperparameter because it controls the complexity, and possibly expressivity, of the network. I thought that a network with larger h may be more capable of abstract reasoning. The final test accuracy values for $K = 1, N = 3$ for h of $64, 128, 256, 512$ were , respectively. Fig 3 shows the test accuracy over iterations for each h. Contrary to my expectation, a smaller hidden dimension resulted in higher accuracy. In particular, $h = 64$ performed the best. One interpretation of this result is that, since this hidden dimension is responsible for encoding the images and labels of the support set into memory, smaller h may promote more abstraction, more informative for the meta-test task.

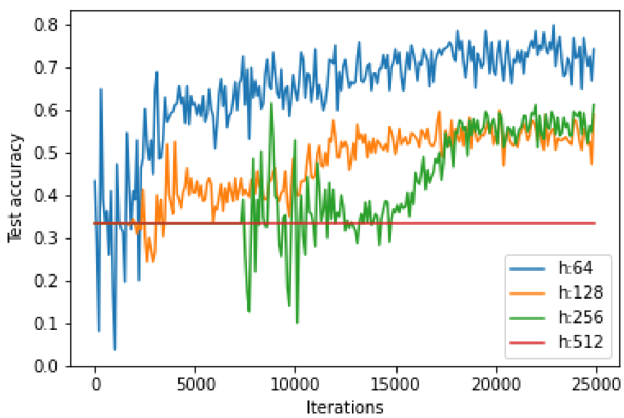

Figure 3: Test accuracy over training iterations for $K = 1, N = 3$ as h is varied

Similarly, Fig 4 overlays the test loss over training iterations. The flat loss curve for $h = 512$ corresponding to a random-initializing accuracy of $1/3$ (note $N = 3$) suggests that optimization is compromised. The same happens for $h = 256$ for the first ~ 7000 steps as well, i.e. green accuracy curve is right on top of the red.

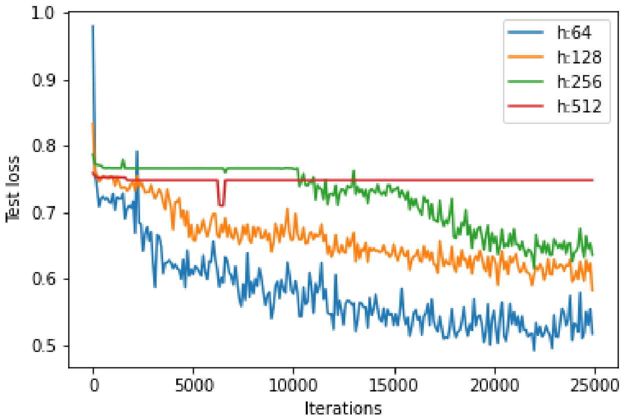

Figure 4: Test loss over training iterations for $K = 1, N = 3$ as h is varied