This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

1.a

In the previous homework recall that we used an LSTM model and that we needed to shuffle the query set so that the model didn't learn to output the same sequence of classes without learning to recognize the images.

For protonets a mapping function is learned that places images of the same class close to each other in feature space. For each class a prototype is calculated. The mapping function is used on the query example placing it in the feature space. Then using a distance function the query example is placed into the class who's class prototype is closest to it. Since then operations of mapping the image into feature space and calculating the distance of the example to the prototype do not use the order of the queries, shuffling the queries is not needed.

1.c.i
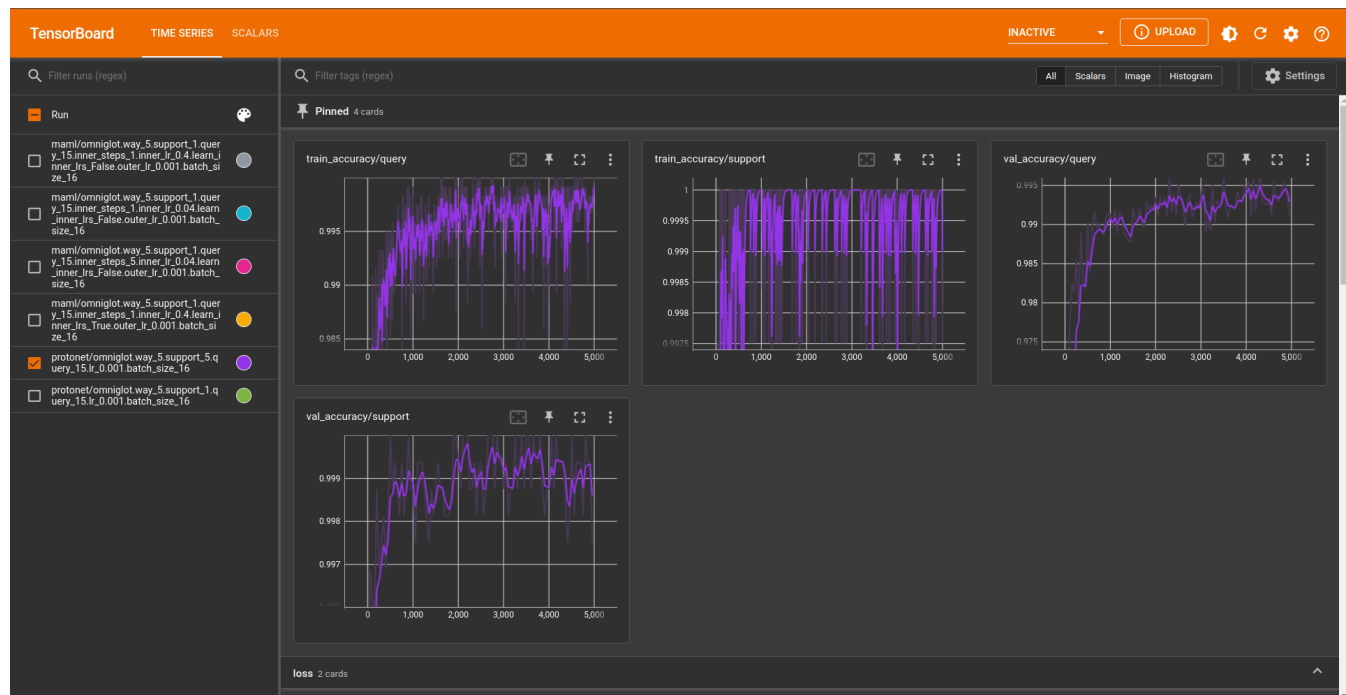
Looking at the graphs above we can see that the train accuracy and validation accuracy for the support classes are around 99%. This indicates that the examples from the same class are being placed close together in feature space.

1.c.ii

Looking at the graphs in 1.c.i it seems that the model is generalizing to new tasks as the validation accuracy for the query and support is around 99%.

1.d.i

	5 way 5 support	5 way 1 support
Mean	0.972	0.98
95% Confidence Interval	0.003	0.002

1.d.ii

The checkpoint of 4900 was chosen because it was the latest checkpoint.

1.d.iii

No there was not a significant difference between 5 shot and 1 shot performance.

To explain why we need to look at what the support features are being used for.

The support features are used for calculating the class prototype which is the mean of the support features.

With a shot of 1 the mean of the support features is the one support feature itself and if it accurately represents the underlying class is as good as the class prototype produced with a higher shot.

What you get with a higher shot is that the class prototype is more likely to represent the underlying class' distribution accurately.
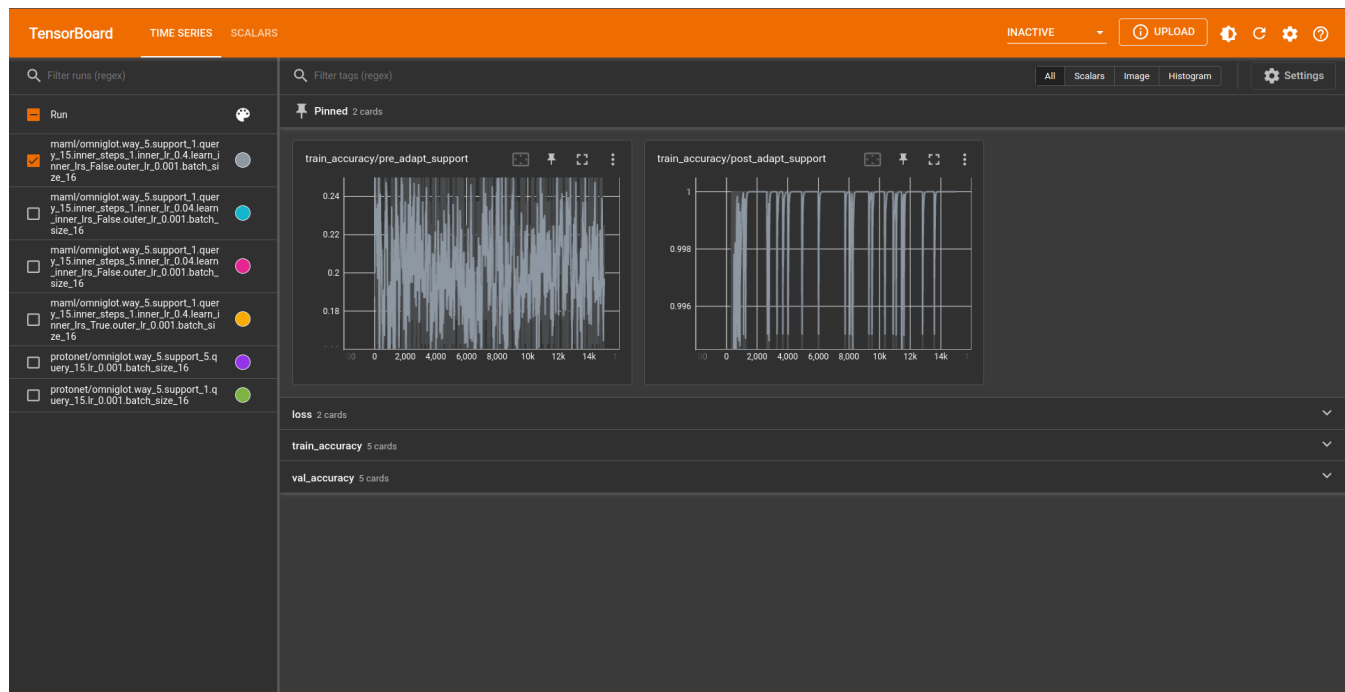
2.c.i
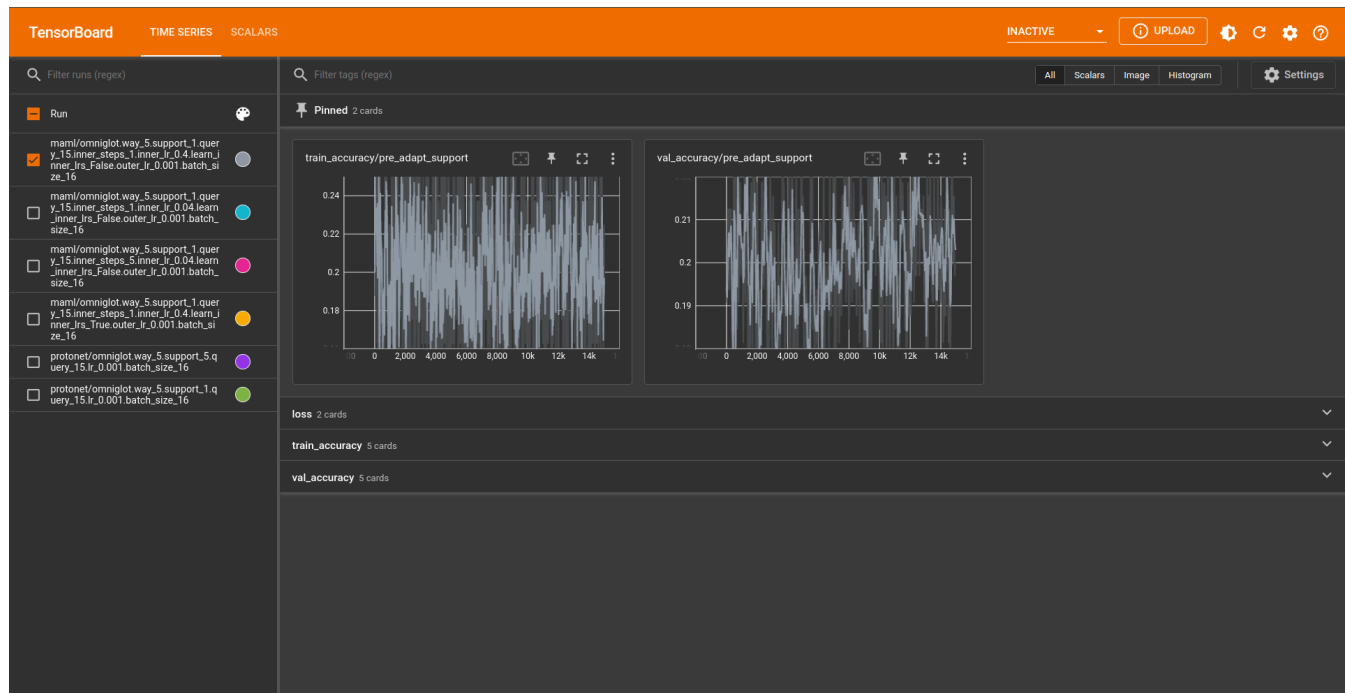
The train and validation accuracy is poor.
This is because the adaptations for each task has not occured yet.
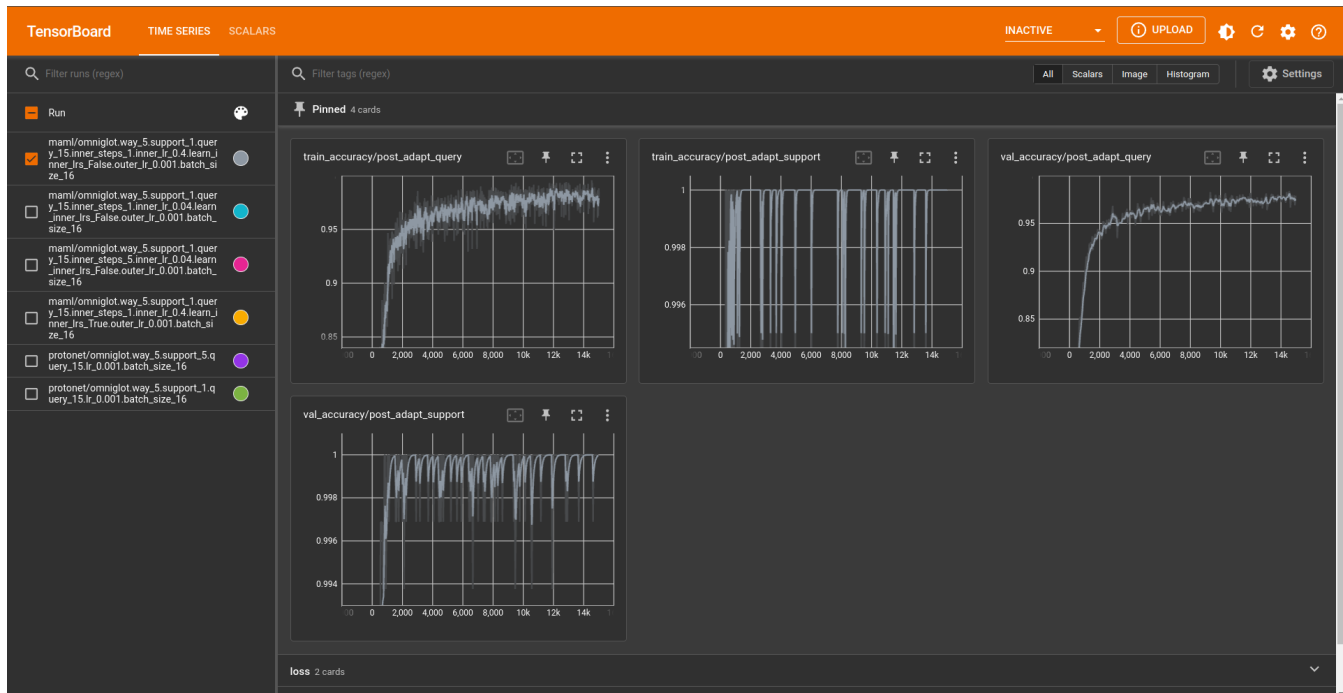The model is fit to all tasks.

2.c.ii

The post adapt support accuracy is much better than the pre adapt support accuracy.
This tells us that the model is not good at any of the particular tasks on it's own without adaptation.

2.c.iii

Both the post train adapt query and train post support accuracies are good with greater than 95% accuracy.
This tells us that the adaptation is working and generalizing to new cases of the tasks.
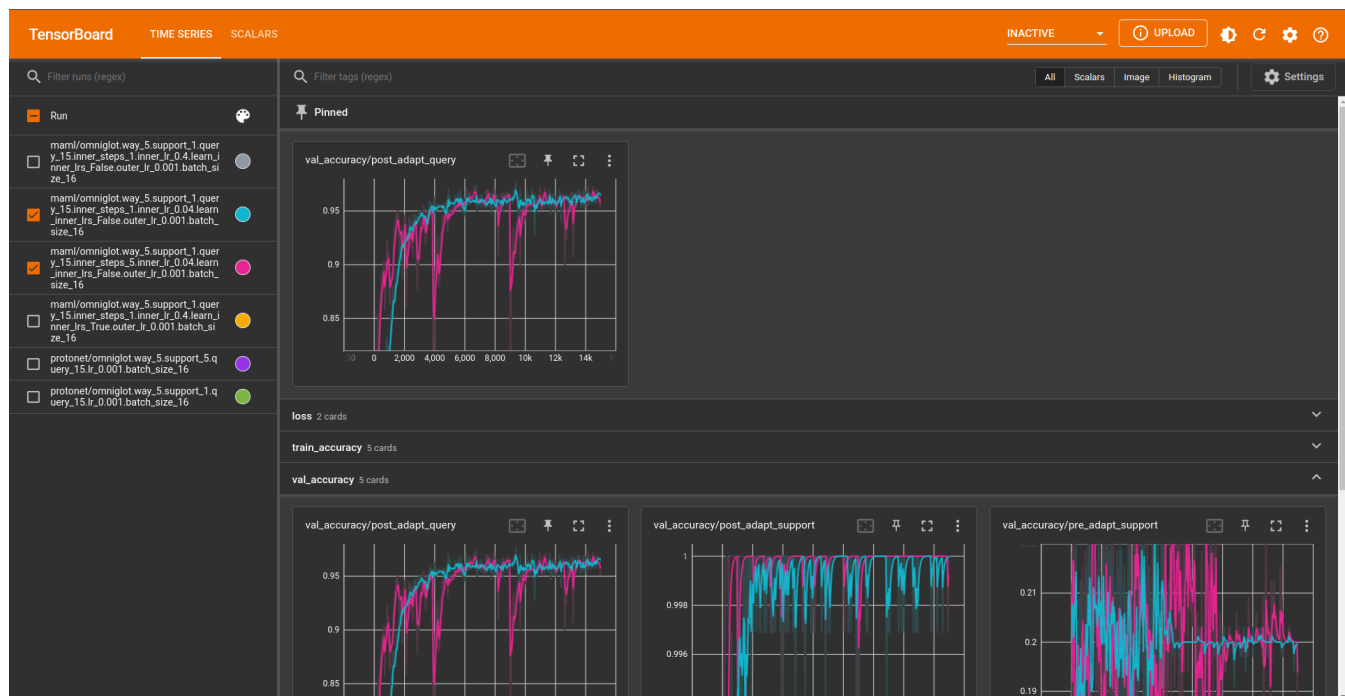Both the post validation adapt query and support accuracies are good with greater than 95% accuracy.
This tells us that the adaptation is working and generalizing to new cases of the tasks.

2.d.i
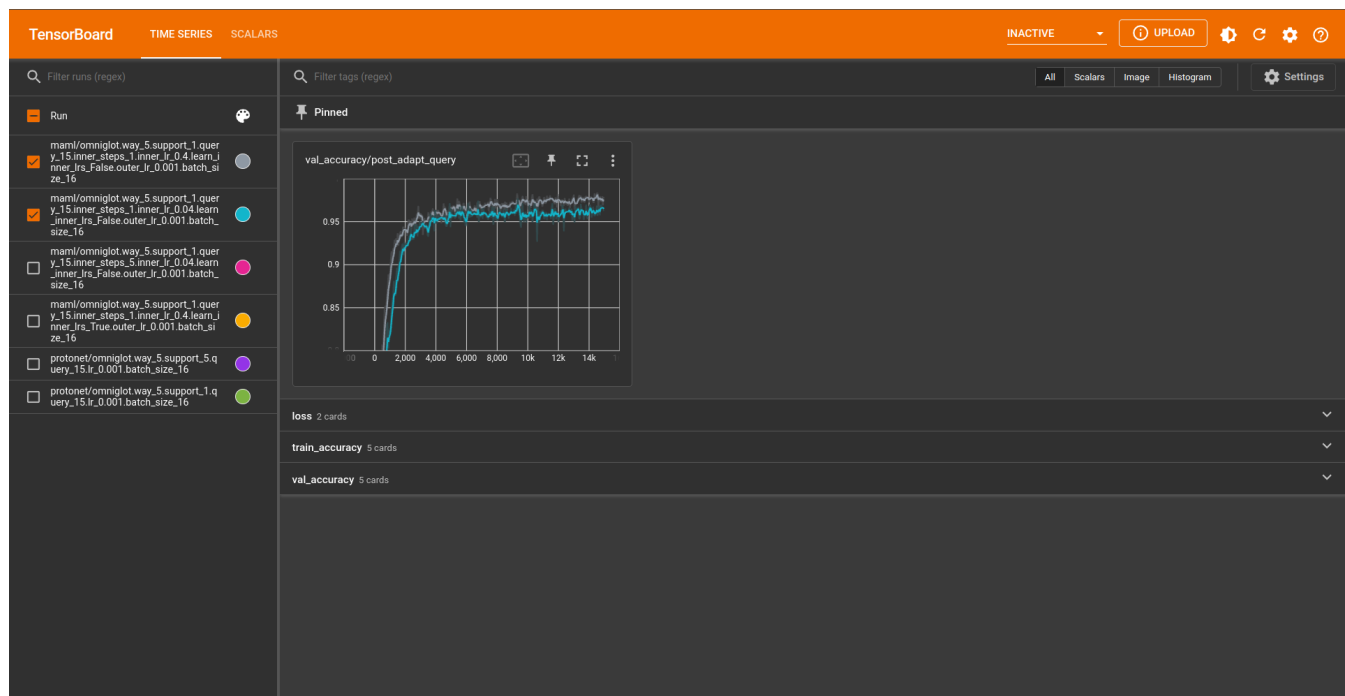
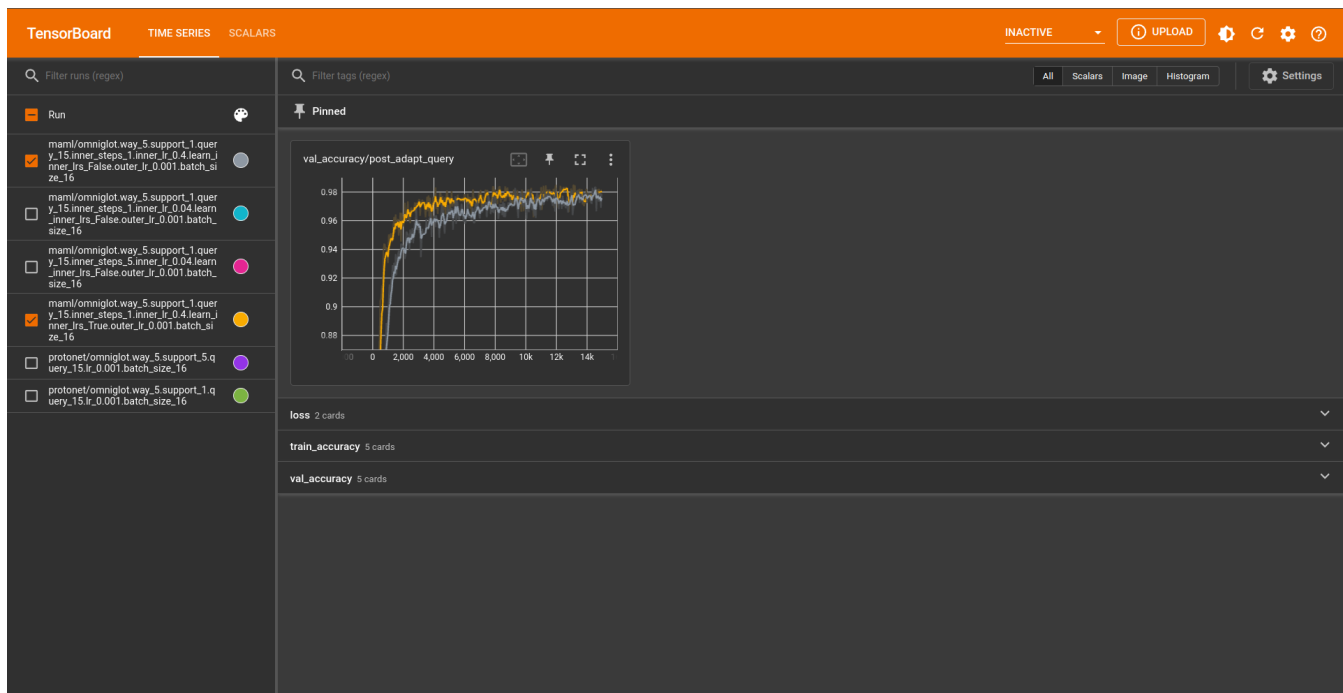Lowering the learning rate makes the model learn slower.

2.d.ii

The effect of increasing the number of inner loop steps is that the model learned faster but took longer to stabilize accuracy.

2.d.iii

The effect of learning the inner learning rates on optimization is that the model learns faster and remains stable. This indicates that outer-loop optimization is faster and continues to generalize.