Concepts and Technologies of Artificial Intelligence

Assignment 2: Random Search Optimisation and Meta Learning

Vadim Stasiev - 1900872

2021

# Introduction

Talk about what was done in the first assignment?

Purpose

Requirements

~~“The success of a machine learning project is often crucially dependent on the choice of good hyperparameters. As machine learning continues to mature as a field, relying on trial and error to find good values for these parameters (also known as “grad student descent”) simply doesn’t scale.”~~

The success of a machine learning solution often depends heavily on the choice of good hyper-parameters.

What’s covered?

This report addresses the objectives of the second assignment which serve as an improvement to the solution developed in the first, more precisely, the brief outlines the need to use one or more strategies to explore the different combinations of hyper parameters, ideally attain one that yields better results. The results that are being sought are ones that surpass the previously developed recognition solution in at least efficiency if not accuracy, given that the accuracy was already high.

# Designing a Solution

As mentioned above, for this assignment it is required the use of at least one of the following suggested strategies:

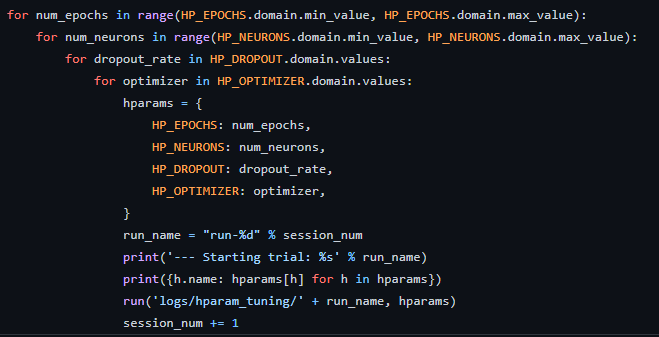
* Random Search
  + Random search is a technique through which random hyper parameter values are tested using a uniform random parameter generator, which ensures that random values are generated uniformly which ensures that different values are tested. (Bengio & Bergstra, 2011)
* Meta Learning
  + In this context, this is referring to the use of an existing model that was trained to optimize the hyper parameters of many different models to optimize the current model. (Hospedales, et al., 2020)
* Adaptive Boosting
  + Also known as “AdaBoost”, it is a machine learning technique that aims to fit models with repeatedly modified versions of the data where the data modifications are made to increase the chances of the model correctly classifying previously incorrect classifications. (sci-kit learn, 2020)
* Cascade Correlation
  + This is a supervised learning algorithm that instead of using a fixed topology, it starts off with a minimal network and automatically trains and adds new hidden layers one by one. This allows the network to learn quickly as well as determine its own size and topology. (Fahlman & Lebiere, 1993)

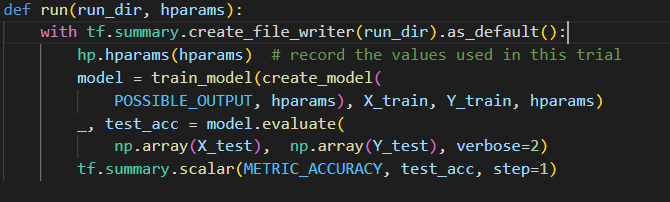
The original solution used an exhaustive search, also known as “grid search”, this was previously used to find the configuration with the best performing number of epochs and neurons, however, for this assignment, it was decided to also add a dropout layer and use its value as another hyper parameter as well as adding the ‘sgd’ and ‘nadam’ as alternative values for the optimizer, which would also be tested individually during the hyper parameter testing.

To start with, it was decided to rewrite the grid search to also include the aforementioned added parameters, to recap on what that is and does, grid search is the process through which it is attempted every single combination of the different hyper parameters. It works iteratively through all of the different combinations, making this the ideal, yet unpractical, solution for finding the absolute best hyper parameter combination.

This time around the implementation also revolved on using a tool provided by “TensorFlow” named “TensorBoard”. (TensorFlow, n.d.) This tool provides a dashboard for viewing all the results of all of the different runs for easy and fast comparison, however the code had to also be adapted to generate compatible data that can actually be interpreted by this tool.

Bellow it is shown the reimplementation of the grid search to generate the mentioned compatible data for interpretation by “TensorBoard”.

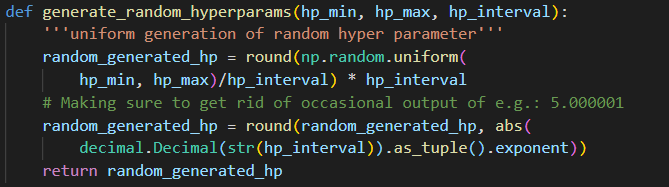




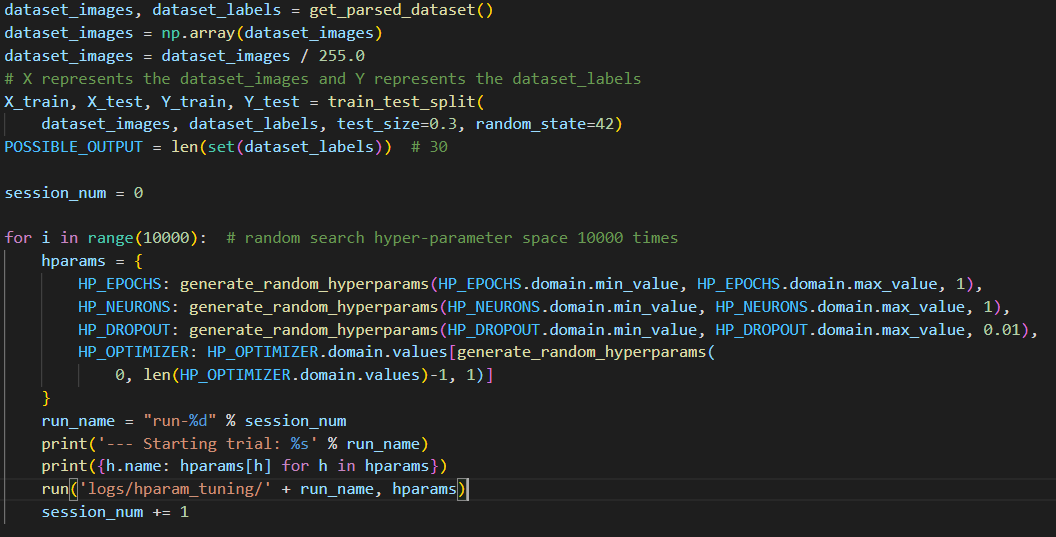
Above it can be seen that all the parameters are being generated and ran iteratively, the function “run” creates the model with the respective parameters and records all the relevant information to the disk.

As already mentioned, the grid search technique is unpractical and this time around due to the addition of the previously mentioned hyper parameters the run took too long so it was eventually cancelled, this also proved the point as to why this technique has hardly any real-world application since any modern machine learning solution, that serves any kind of purpose, would have at least this level of complexity as far as hyper parameters go.

Moving on to using at least one of the suggested strategies, for this assignment, it was decided to use “Random Search” since the implementation of it seemed fairly straight and promising as long as the parameter values would be uniformly generated such as to ensure no single value was tested more than the others. In order to do this, the following function was created:



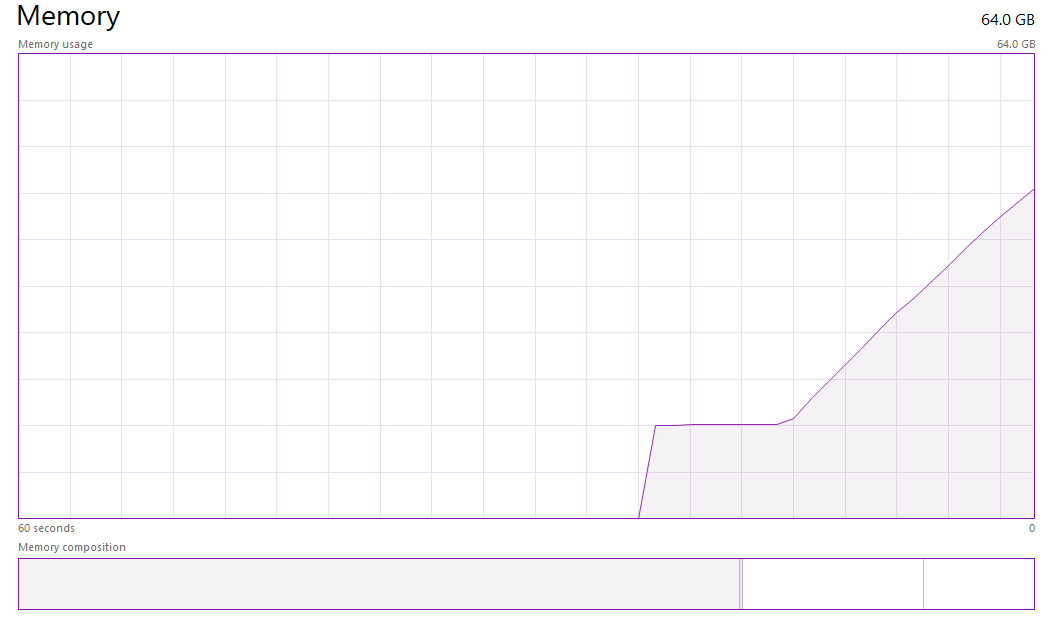
The function was then implemented in the following way:



As it can be seen above, there is a for loop that runs 10001 times and over each run different random hyper-parameters are generated uniformly, the run results are then also saved for later analysis.

## Unexpected Issues

Despite looking like a promising tool, TensorBoard appears to have a memory leak which has been present in the software for a few years at least (TensorFlow GitHub, 2021), and has not been fixed yet, this is seems to be triggered whenever TensorBoard has to load a lot of run logs, the number of logs tested in this scenario was roughly 10000 and it seemed to fill up the entire memory of the system in the space of a few seconds, naturally if not stopped the memory leak would not stop there and would continue onto to windows paging file. This issue might be related to the windows NTFS file system but that was not pinpointed, it was instead found a work-around, to work around the issue the “TensorBoard” software needs to be started on an empty directory and then after the fact, move the logs into the working directory.



# Experiments

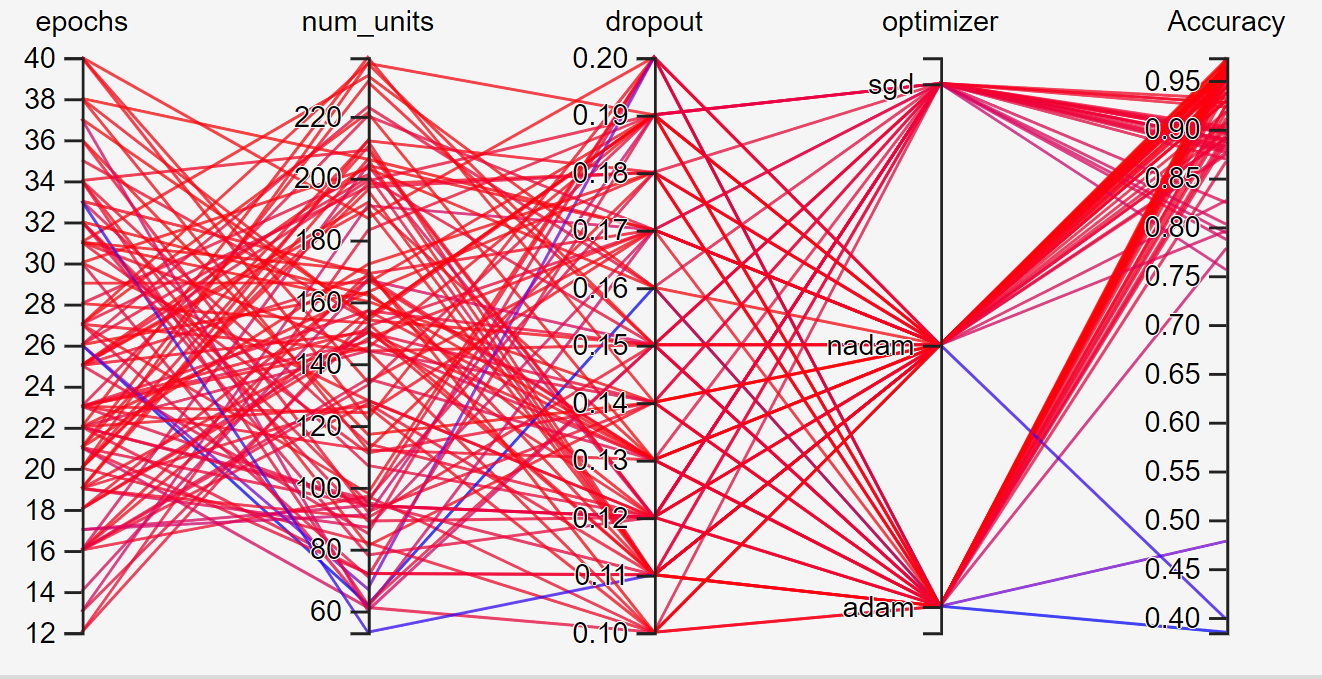
As previously mentioned, the experiments conducted using grid search were cancelled so they will not be included.

As for the random search experiment, this was concluded with 10,000 runs with the following tested values per parameter:

|  |  |
| --- | --- |
| Epochs | From 12 to 40 |
| Neurons | From 50 to 240 |
| Dropout Rate | 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.17, 0.18, 0.19, 0.2 |
| Optimizer | 'adam', 'sgd', 'nadam' |

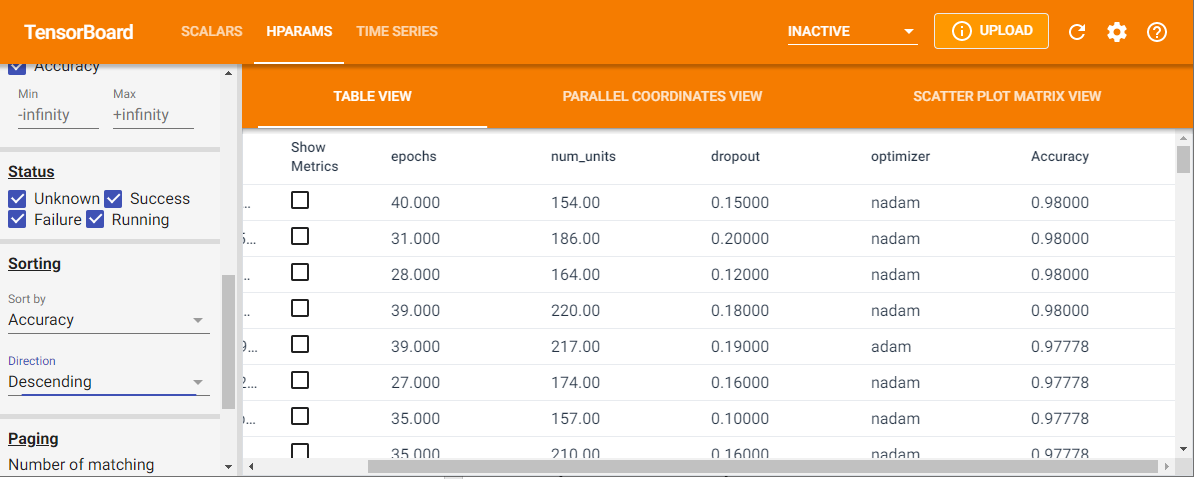
As these same values were used in the original grid search, math would suggest that it would have resulted in a staggering 238,260‬ runs, so while 10,000 runs may seem much, it actually equates to roughly only 4% of the total number of combination possibilities.

The resulting graph of all the runs can be seen bellow:



While this graph can give a rough impression of how the different parameters affected the model’s accuracy, it looks fairly dense and therefore it is hard to drawn actual results from it, to do that we must instead look at the top results and compare them to each other.

On the TensorBoard, with all the runs selected, under the HPARAMS tab, there is some functionality that allows the data to be sorted, in order to see what the best results were, this data needs to be sorted by the accuracy in a descending direction.



On the resulting table it can be seen the best results of all of the 10,000 experimental runs and while it may not always be the case, in this case the top result seems to not only have the same accuracy as the few before it, but it also does so by being the most efficient one as it has the least number of hidden neurons, hitting an accuracy of 98% with only 154 neurons.

# Conclusion