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:

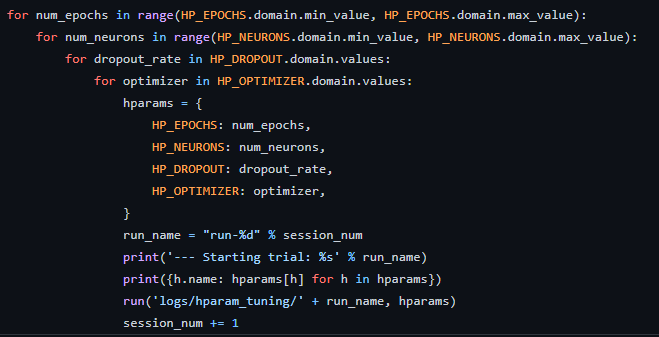
* Random Search
  + Random search is a technique through which random hyper parameter values are tested using a uniform random parameter generator, which ensures that random values are generated uniformly which ensures that different values are tested. (Bengio & Bergstra, 2011)
* Meta Learning
  + In 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”.



As already mentioned, the grid search technique is unpractical and this time around due to the addition of a few hyper parameters, in comparison to the solution developed in the first assignment, 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 solutions that serves any kind of purpose would have at least this level of complexity as far as hyper parameters go.