Deep Artificial Neural Network Optimization

Computational Intelligence for Optimization
Course Project

Illya Bakurov
email ibakurov@novaims.unl.pt

Leonardo Vanneschi
email lvanneschi@novaims.unl.pt

07/11/2018

Abstract

The objective of this document is to introduce students to the project of Computational Intelligence for Optimization curricular unit, and provide them with all necessary information for its successful completion. The objective of the project is to optimize 64000 weights in a predefined Deep Artificial Neural Network architecture, which task is to classify handwritten digits. To accomplish this task, students must use optimization algorithms aborted during the classes.

Introduction

Human Learning

Our brain is an incredible supercomputer that never stops working, unless it dies. This powerful tool allowed us, Humans, to survive and improve throughout thousands years of evolution. It made us unique and mighty, relatively to the other species. All the best we can do - talk, draw, play music, build cities, boats, computers, planes, space rockets, etc. - is due to our brain.

The previously listed abilities, among many others, were made possible thanks to an even more powerful one - the ability to learn. According to IGI Global [1], *completion of the learning cycle (...) includes active testing, concrete experiences, reflective observation, and abstract hypothesis*. Considering the academic context, students learn by means of examples which are shown during the classes (the *training data*). The examples are followed by solutions (also called *labels* or *targets*) which are necessary to make them understand their mistakes.

One way to evaluate students ability to learn is to show them new, previously *unseen* examples (also called the *test data*), although not very different from the ones they used to study before. Usually, this procedure happens during the intermediate *test* and exams.

What distinguishes a good student from a bad student is the *generalization ability*. A bad student memorizes those examples he/she studied before the exam, and, consequently, fails during the testing procedure. A good student understands the underlying knowledge, hidden in the training data, and is able to generalize it to new, previously unseen, examples.

Machine Learning

Machine Learning (ML) is a vast sub-field in Computer Science which main objective is to provide computers (machines) the ability to learn. One way to accomplish this task consists of simulating the human learning.

Supervised Machine Learning

Supervised ML (SML) aims to infer a relationship, mathematically materialized as a function, provided a set of training data, which holds the underlying relationship between input features and the targets. One of the main concerns of this input/output mapping is the generalization ability of the machine, i.e., the ability to discover general and persistent patterns between inputs and targets in the training data that are possible to generalize to another, previously unseen, data. Depending on level of measurement of the target, the supervised learning task can be defined as regression (for continuous target values) or classification (for discrete target values).

Deep Learning

The term Deep Learning (DL) is frequently used to entitle an Artificial Neural Network (ANN) with *many*, usually more than one, hidden layers. An ANN itself is a biologically-inspired computer system which simulates the Biological Neural Network (BNN) and its biochemical processes. As a result, an ANN consists of a set of interconnected layers of neurons (basic processing units) which, all together, form a powerful and versatile computer system, able to solve numerous complex optimization problems (OPs) like speech recognition, image classification, natural language processing, drug discovery, fraud-detection, autonomous car driving, etc.

The *intelligence* in an ANN system is embedded in the way as neurons from separate layers are interconnected and the relative strength of different connections.

The Project

The course project consists of finding an optimal set of weights for 64000 connections in a predefined Deep Artificial Neural Network (DANN) architecture, which task is to (correctly) classify handwritten digits. In [2] one can find additional information regarding the data that is used to process the network.

The DANN

Figure 1 provides a visual representation of the DANN architecture to be considered during this project:

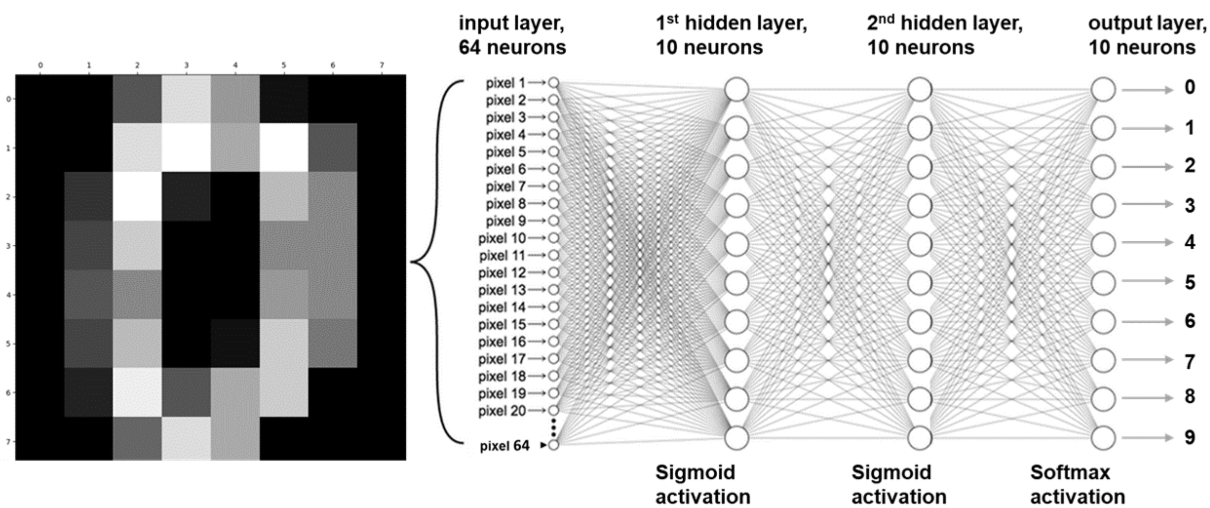

Figure 1: 64x10x10x10 DANN architecture for a 10-label classification problem.

The inputs of the network are 8x8 raster images of handwritten digits. For every pixel in the image there is a corresponding integer value in [0, 16] (gray scale). The images are flattened and provided to the network one image at a time. For this reason, the first (input) layer of the network consists of 64 neurons.

The network is composed by 2 hidden layers, 10 neurons each. The neurons in both layers use a sigmoid activation function [3]. Given that the SML task for this ANN consists of a 10 label classification, the output layer is composed by 10 neurons - one for each output label (or class) - and the activation function is Softmax [4].

Taking in consideration the 64 input neurons, the 10 neurons in each hidden layer, which are 2, and the 10 output neurons, the network has 64000 connections.

DANN Optimization

The Solution

Given that the primary goal is to find an optimal set of weights for connections of the network in figure 1, a candidate solution is, at a technical level, a *numpy.ndarray* of 64000 floating point values.

The Optimization Problem

The problem of finding a set of weights for an ANN, regardless its architecture and task, is, in fact, a special type of continuous OP. Conceptually, the problem is not that different from the one of optimizing a Sphere [5] cost (a.k.a. fitness) function. The main difference consists in the fitness evaluation of candidate solutions - the solutions are evaluated first and only then validated. The validity of a solution depends on the absolute difference between training and validation accuracy - if there is no *significant* difference, then it is valid.

Fitness Function

In the context of ANN OP, the fitness function is a concrete ANN. In the context of the course project, the DANN presented in figure 1 evaluates a given candidate solution by taking its representation - a set of 64000 weights - and classifying the handwritten digits. The fitness of a candidate is calculated as network's classification accuracy on training and validation data.

Objectives and Evaluation Rules

Objectives

Taking in account the optimization algorithms studied throughout the course, consider the following list of main objectives when solving the project:

1. **explore**: benchmark different optimization algorithms and their parameters. It is recommended to use at least 5 runs to generate summary measures;

2. **exploit**: research, implement and benchmark new variations of optimization algorithms;

3. **focus**: focus of your project should be on Genetic Algorithms and their aspects.

It is important to notice that all the benchmarks should be performed under fair conditions, i.e., under equivalent computational effort. Also, specially regarding second objective, include all necessary bibliographic references.

Evaluation Rules

The evaluation of the student i in group g will based on the following formula:

$$grade_i = .1 * report_g + .1 * presentation_i + .4 * unseenFitness_g + .4 * techniques_g,$$

where *techniques* stands for the diversity, complexity and originality of applied techniques.

The following is a list of mandatory requirements. Violation of at least one of them, will lead to reprobation of all the elements the group:

- changing of the architecture of DANN;
- using more than 5000 fitness evaluations per run, for a given instance of a search algorithm;
- using more than 50 fitness evaluations per generation, for a given instance of a search algorithm;
- changing the training data;
- changing the proportion of test data (it should be 33%;
- changing the proportion of validation data (it should be 20% from the remaining 77%;
- changing the validation threshold (it should be 7%);
- changing the configuration of search space in ANN OP (weights should be initialized in $[-2,\ 2[$.

References

[1] Igi global international publisher of information science and technology research. https://www.igi-global.com/dictionary/analyzing-farmers-learning-process-in-sustainable-development/16939.html.

[2] Mnist database: optical recognition of handwritten digits dataset, sklearn. https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load$_{d}igits.html$.

[3] Sigmoid function. https://en.wikipedia.org/wiki/Sigmoid$_{f}unction.html$.

[4] Softmax function. https://en.wikipedia.org/wiki/Softmax$_{f}unction$.

[5] Sphere cost function. https://www.sfu.ca/ ssurjano/spheref.html.