# DARTS

Henk Jekel (5609593)

April 28, 2023

# 1 DARTS

## 1.1 Motivation for NAS

Deep learning can be very powerful, but different problems require different neural architiectures (no one size fits all architiecture). In order to fit the architecture to a specific problem, hyperparameter search is used in which answers to the following questions are answered:

- How many layers?
- How many nodes per layer?
- How to connect different layers?

One way to find a good neural architecture for a specific problem automatically is neural architecture search. The following section describes a specific type of neural architecture search that uses gradient descent to find the optimal hyperparameters.

## 1.2 Darts explained

Differentiable architecture search (DARTS) aims to do neural architecture search (NAS) using gradient descent. Previously, NAS was performed using reinforcment learning algorithms as NAS involves discreet choises. Examples of such discreet choices are:

- The spatial dimension of a kernel
- The block used at a specific depth in the architecture

As mentioned, naturally NAS is not approached with a gradient based search as the problem involves discreet decisions which are typically not differentiable, see figure 1. The autors of DARTS opted for a continuous architecture

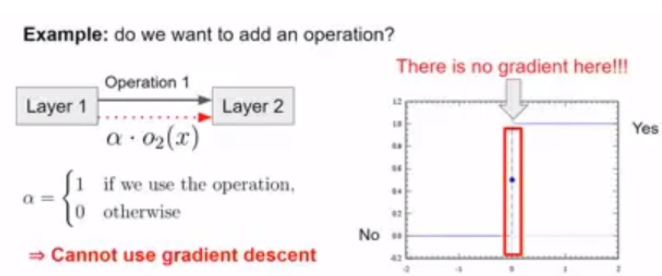


**Figure 1: Non-differentiable search space of hyperparameter search**

coefficient $\alpha$ mentioned in figure 1 varying anywhere between 0 and 1, see figure 2.

To apply gradient decent which is used for continuous spaces, the authors came up with a search space containing cells. Here, a cell is a directed acyclic graph consisting of an ordered sequence of N nodes. Acyclic means that the graph does not contain any cycles and directed means that the edges in the graph can only be traveled in one direction. Ordered refers to the numbering of the nodes. An example of this graph is displayed in figure 3. In this figure, the autors gave an example of a set of operations that can be used between each set of layers represented in Equation 1.

$$O \in \{1 \times 1, 3 \times 3, 5 \times 5\} \quad (1)$$

The numbered nodes $x^{(i)}$ refer to the output feature maps in the architecture. The edges represent operations $o^{(i,j)}$ transforming node $i$ into node $j$, such as fully connected layer, convolutional layer etc..

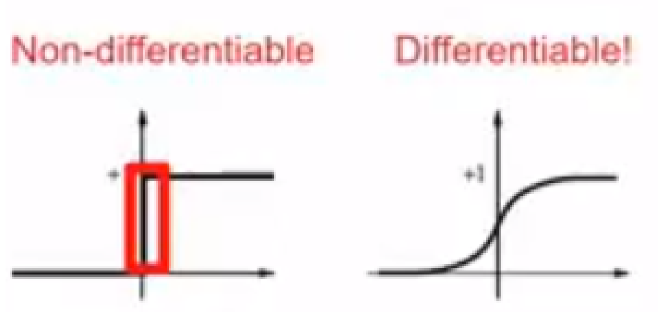


**Figure 2: Continuous architecture coefficient $\alpha$**

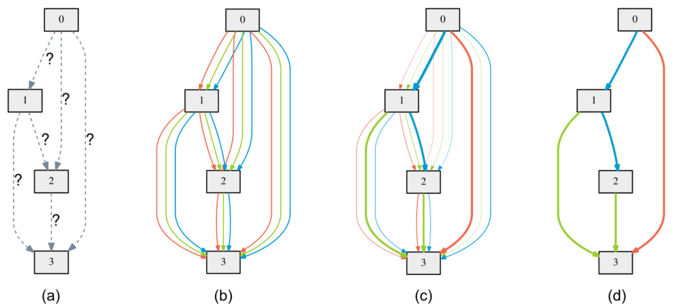


**Figure 3: Graph of a cell used in DARTS**

## 1.3 DARTS search

In the context of DARTS (Differentiable Architecture Search), "normal" and "reduce" refer to two types of cells that are used in the architecture search process.

A "normal" cell is a cell that maintains the same spatial resolution between the input and output, while a "reduce" cell is a cell that reduces the spatial resolution between the input and output.

Specifically, in DARTS, each cell is constructed by stacking multiple normal and reduce cells in a predefined pattern. During the architecture search process, the weights of each cell are learned by backpropagation, and the normal and reduce cells are chosen by a learnable weight vector.