Optimization for training Deep Models

- **Faizan Shaikh**
Trainee Data Scientist
Analytics Vidhya

Topics Covered in the Book

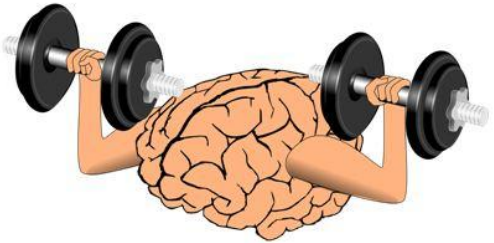
- Optimization in DL vs Pure Optimization
- Challenges in optimizing neural networks
- Optimization techniques - (Basic and advanced)
- Optimization strategies

Quick Intermission

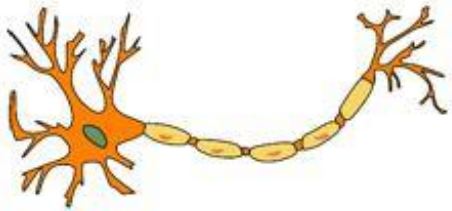
What have we covered until now?

Review Questions

Q1. A neural network model is said to be inspired from the human brain.



The neural network consists of many neurons, each neuron takes an input, processes it and gives an output. Here's a diagrammatic representation of a real neuron.



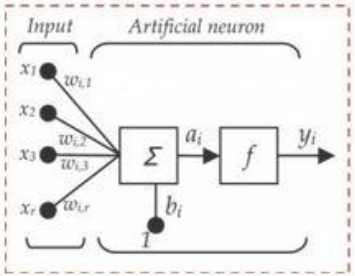
Which of the following statement(s) correctly represents a real neuron?

A. A neuron has a single input and a single output only
B. A neuron has multiple inputs and multiple outputs
C. Both of the above statements are true

Reference - https://www.analyticsvidhya.com/blog/2017/01/must-know-questions-deep-learning/

Review Questions

Q2. Below is a mathematical representation of a neuron.



The different components of the neuron are denoted as:

- **x1, x2,…, xN: These are inputs to the neuron. These can either be the actual observations from input layer or an intermediate value from one of the hidden layers.**
- **w1, w2,…,wN: The Weight of each input.**
- **bi: Is termed as Bias units. These are constant values added to the input of the activation function corresponding to each weight. It works similar to an intercept term.**
- **a: Is termed as the activation of the neuron which can be represented as**
- **and y: is the output of the neuron**

$$a = f(\sum_{i=0}^{N} w_i x_i)$$

Considering the above notations, will a line equation (y = mx + c) fall into the category of a neuron - Yes or no?

Reference - https://www.analyticsvidhya.com/blog/2017/01/must-know-questions-deep-learning/

Optimization: Introduction

Optimization: Introduction

Given neural network parameters $\boldsymbol{\theta}$, find the value of θ that minimizes cost function

$$J(\theta).$$

-Exhaustive search
-Random search (genetic algorithms)
-Analytical solution
-Model-based search (e.g. Bayesian optimization)
-Neural nets usually use **gradient-based search**

Reference: http://www.deeplearningbook.org/slides/sgd_and_cost_structure.pdf

Pure Optimization vs Optimization in DL

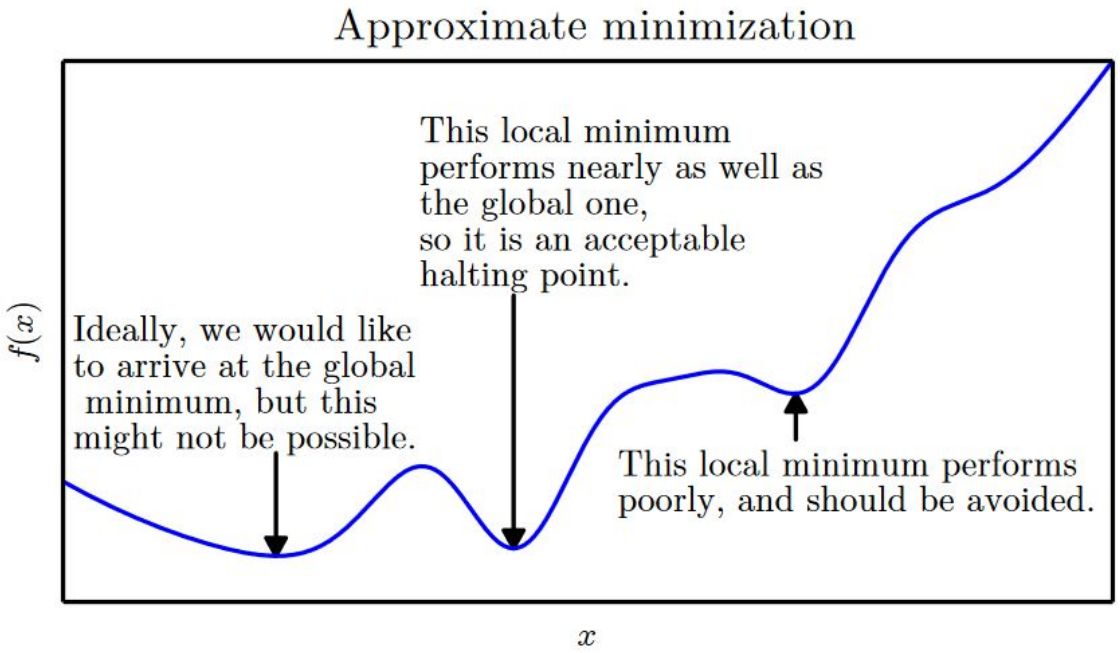
Example 1:

```
#pragma optimize("", off)
// some code here
#pragma optimize("", on)
```

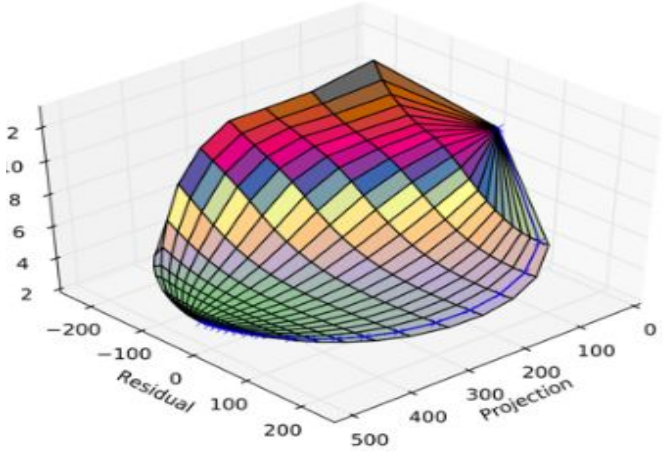
Example 2:

https://www.analyticsvidhya.com/blog/2016/09/a-beginners-guide-to-shelf-space-optimization-using-linear-programming/

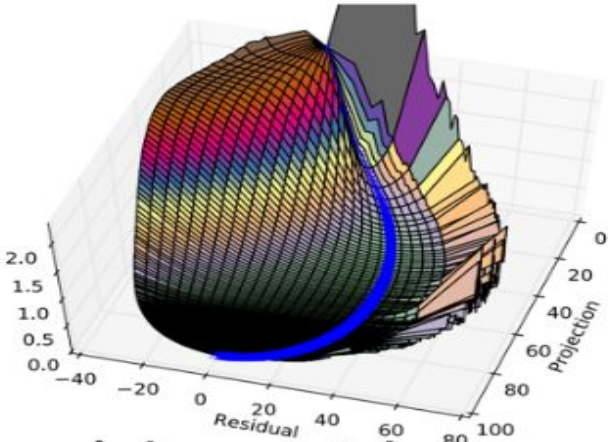
Challenges in Optimization



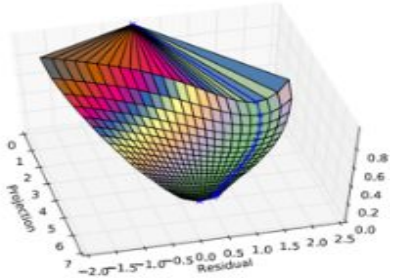
Challenges in Optimization



LSTM



Adversarial ReLUs



Factored Linear

(Goodfellow 2015)

Optimization Techniques

Most important

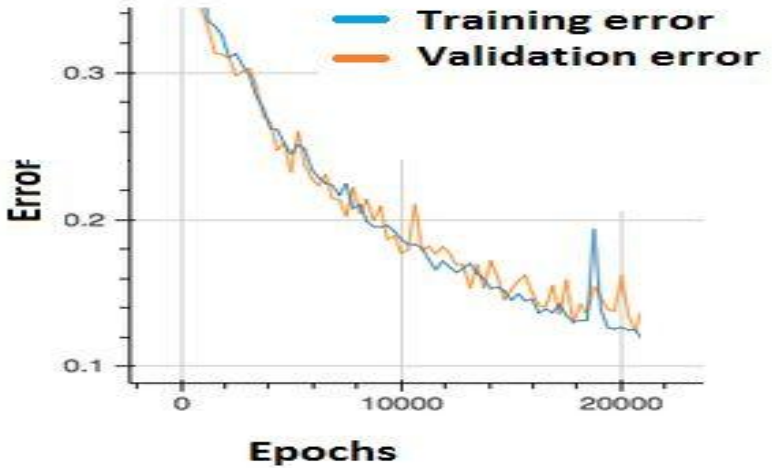
- SGD*
- SGD with Momentum*
- Adam**

* - Basic

** - Adaptive learning

Quick Question

In the graph below, we observe that the error has many “ups and downs”



Should we be worried?

A. Yes, because this means there is a problem with the learning rate of neural network.

B. No, as long as there is a cumulative decrease in both training and validation error, we don’t need to worry.

Thank you!

Notes:

- What is optim
- Diff betn pure optim and DL optim
- Challenges
- Basic algorithms
- Advanced algos
- Strategies