|  |  |  |
| --- | --- | --- |
| Ain Shams University  Faculty of Engineering  Mechatronics Department | Jobs and Careers at Ain Shams University in Egypt – Join Us Today! | Computational Intelligence  CSE 473s  Computer and System Department |

**Major Tast Project first milestone :**

**Performance of different Optimization Algorithms**

|  |  |
| --- | --- |
| Name | ID |
| Hussein Hamed Hussein | 2001767 |
| Mostafa Mahgoub Sobhy | 2001131 |
| Ahmad Radwan Farouq | 2000114 |

Submitted to:

Prof.Hossam Hassan

Eng. Abdulah Awadallah

**Abstract**

Optimization algorithms are essential components in training neural networks, significantly affecting their learning efficiency and final performance.

This study examines the performance of several widely used optimization techniques, including.

* SGDR
* NAG
* RMSProp
* Nadam
* Learning Rate Schedulers
* Adam
* Momentum

A standard dataset and a consistent shallow neural network architecture are used to ensure a fair comparison, with the evaluation focusing on critical metrics such as training speed, convergence behavior, and accuracy.

The analysis highlights the trade-offs between computational efficiency and model effectiveness for each optimization strategy, offering a balanced perspective on their applicability.

The results of this work provide valuable insights into the selection of optimization algorithms, contributing to informed decision-making in a wide range of machine learning applications.

**Problem Definition**

In the following context, let us demonstrate the problem by focusing on three main points.

Goal: Compare the performance of different optimization algorithms for training a shallow neural network on a 10 classes data set

* CIAFAR 10
* MNIST

Metrics: evaluating the model performance using the following aspects

* accuracy,
* training time
* potential stability issues.

Tasks: Implement the different optimization algorithms to be able to select the suitable candidate for the problem

Such as

* Stochastic Gradient Descent with Warm Restarts
* Nesterov Accelerated Gradient (NAG)
* RMSProp
* Nadam
* Learning Rate Schedulers (Exponential Decay, Step Decay)

**Dataset Preparation**

The following step in data preparation for our formulated problem

(classification problem ) involves.

* normalizing the pixel values
* labels are converted into one-hot encoding format.
* dataset is split into training, validation, and test sets.

**Neural Network Design**

To design the neural network for classification problems, you must consider three main parts.

* Dimensions of input vector
* Task Complexity
* Dimensions of output vector

For Input layer

* The number of units in the input layer is determined by (i.e., Images) the resolution and number of color channels in the image.

For Hidden layers and Units

* Required performance.
* Available Computational power

For Dimensions of output vector

* The number of units in the output layer is determined by the number of output classes.

Input layer:1024 units for CIFAR 10,784 units for MNIST.

Hidden layers: two layers (shallow network)

Units in the first hidden layer: 512 units.

Units in the second hidden layer: 256 units.

Output layer: ten units for both CIFAR 10 and MIST

**Implementation of Algorithms and Results**

For CIFAR 10

|  |  |  |
| --- | --- | --- |
| Algorithm | Training Time | Accuracy over epochs |
| SGDR | 50s |  |
| NAG | 54s |  |
| RMSprop | 49s |  |
| Nadam | 66s |  |
| Learning Rate Schedulers (Exponential Decay) | 53s |  |
| Learning Rate Schedulers (Step Decay) | 62s |  |
| Benchmark  adam | 52s |  |

For MNIST

|  |  |  |
| --- | --- | --- |
| Algorithm | Training Time | Accuracy over epochs |
| SGDR | 88s |  |
| NAG | 80s |  |
| RMSprop | 85s |  |
| Nadam | 86s |  |
| Learning Rate Schedulers (Exponential Decay) | 84s |  |
| Learning Rate Schedulers (Step Decay) | 80s |  |

**Experimental Results discussions**

Benchmark

|  |  |  |
| --- | --- | --- |
| Architecture | MNIST | CIFAR10 |
| Baseline Models | Simple Neural Network (1-2 layers): ~94-95% | Simple Neural Network (1-2 layers): ~50-60% |
| Classical Machine Learning Techniques | SVM with RBF Kernel: ~97-98% | SVM with RBF Kernel: ~55-65% |
| Deep Learning Models CNNs | Simple CNN (3-4 layers): ~99.4-99.6% | ResNet ( ResNet-50): ~95% |
| State-of-the-Art Results | VGG, ResNet:  ~99.7-99.8% | ResNet DenseNet  ~98% |

Model Best Performance

For CIAFAR 10 achieved 50% accuracy with adam optimization

Conclusions:

* The model meets the benchmark results and could be enhanced with more complected model and advanced architecture.

For MNIST achieved 97% accuracy with RMSprop and Nadam optimization

Conclusions:

* The model meets and exceeds the benchmark result due to proper parameters tunning and data normalization.