2802ICT Intelligent Systems Assignment 2

Michael Ellicott
s5276830

May 25, 2023

Contents

1 Classification using Decision Trees

1.1 Software Design

- Node class
- DecisionTree class

1.2 Experiments

1.3 Findings

2 Classification using Neural Networks

2.1 Software Design

NeuralNetwork Class

- **NeuralNetworks** can be instantiated either through passing the network architecture, learning rate, number of epochs and mini-batch size, or by passing a filepath to a file containing precomputed weights and biases.
- **__forwardpropagate**(input)
  Performs forward propagation using input as the input data.
- **__backpropagate_collect_avg**(target)
  Performs back propagation and accumulates the gradient. This method is called once for every sample in a mini-batch.
- **__backpropagate_apply_avg**(target)
  Averages and applies the gradients accumulated by **__backpropagate_collect_avg** to the weights and biases of the network.

2.2 Experiments

10 different models were trained with various mini-batch sizes and learning rates over 50 epochs in order to investigate the effect of these hyperparameters on model accuracy.

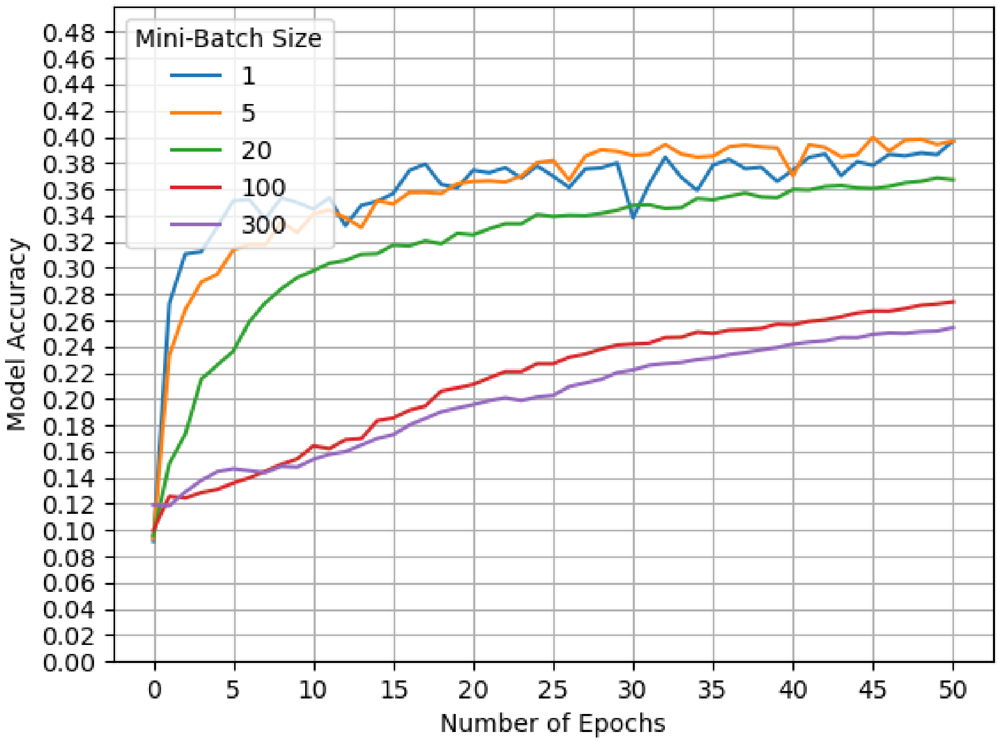

Figure 1: Effect of mini-batch size on accuracy over epochs
learning rate = 0.1

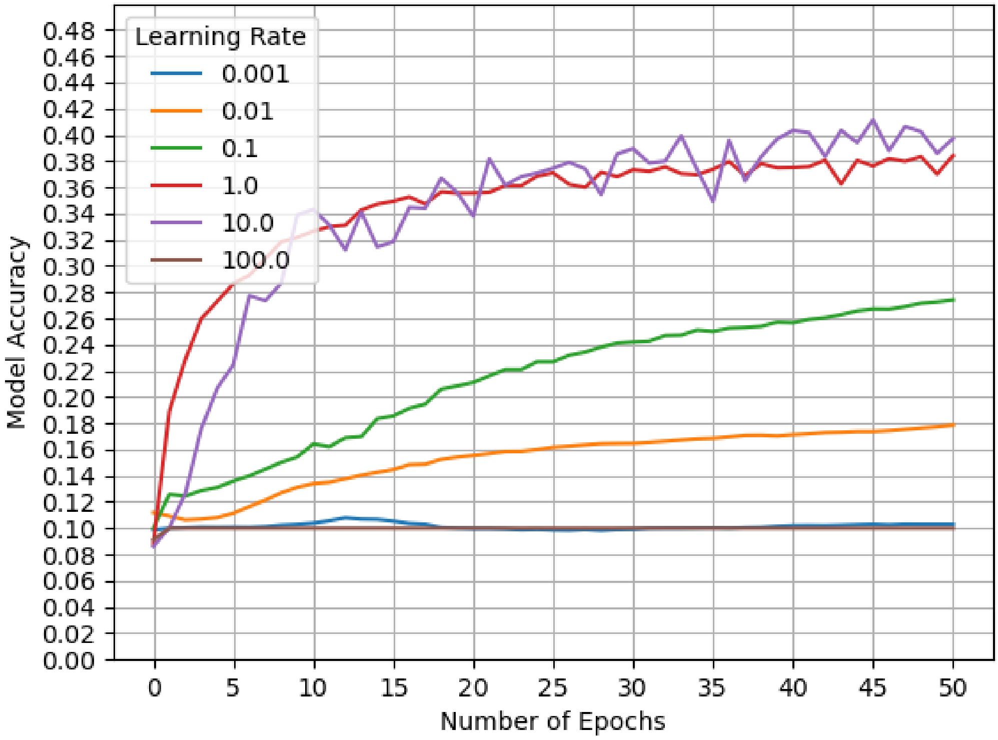

Figure 2: Effect of learning rate on accuracy over epochs
mini-batch size = 100

2.3 Findings

Mini-Batch Size Mini-batch size has a significant effect on how quickly the error is reduced. A lower mini-batch size tends to provide faster, but more erratic error reduction, wheras a higher mini-batch size yields a more gradual reduction in error due to the gradient vector taking more of the overall dataset into account. Interestingly, a high mini-batch size results in any given epoch being much more likely to have a lower error than the epoch preceding it, although at the cost of training time.

One method to take advantage of this information may be to train the network for some number of epochs using a much lower mini-batch size. After which the training can continue using a higher mini-batch size, allowing the total error to reduce quickly at the beggining and then more smoothly over time.

Learning Rate Learning rate also has a great effect on the rate of error reduction over epochs. Its effect is similar to that of mini-batch size, with the major difference being that too high or too low of a learning rate appears to trap the model in a local optimum. Inside the range of learning rates which do not trap the model in a local optimum, a lower learning rate produces a more gradual decrease in error. Likewise, a higher learning rate tends to decrease the error more rapidly but also more erratically.

It may also be possible to take advantage of this, too. By training the network using a high learning rate for some number of epochs and then switching to a lower learning rate, a more accurate model may be reached sooner than by using a lower learning rate alone.

Variable Mini-Batch Size and Learning Rate By taking advantage of the information learned from the above experiments, we make estimate when the optimal time to change both the mini-batch size and learning rate in order to train a model faster.

A mini-batch size of 1 decreases the total error rapidly until approximately epoch 5 (Figure 1). Similarly, a learning rate of 10 decreases the error rapidly until approximately epoch 15 (Figure 2). By taking advantage of this information and varying the mini-batch size and learning rate we can train a model with the same accuracy in much less time. Firstly, we begin training with a mini-batch size of 1 and a learning rate of 10. At epoch 5 we increase the mini-batch size to 20 and at epoch 10 we decrease the learning rate to 0.1.

This yields the following result

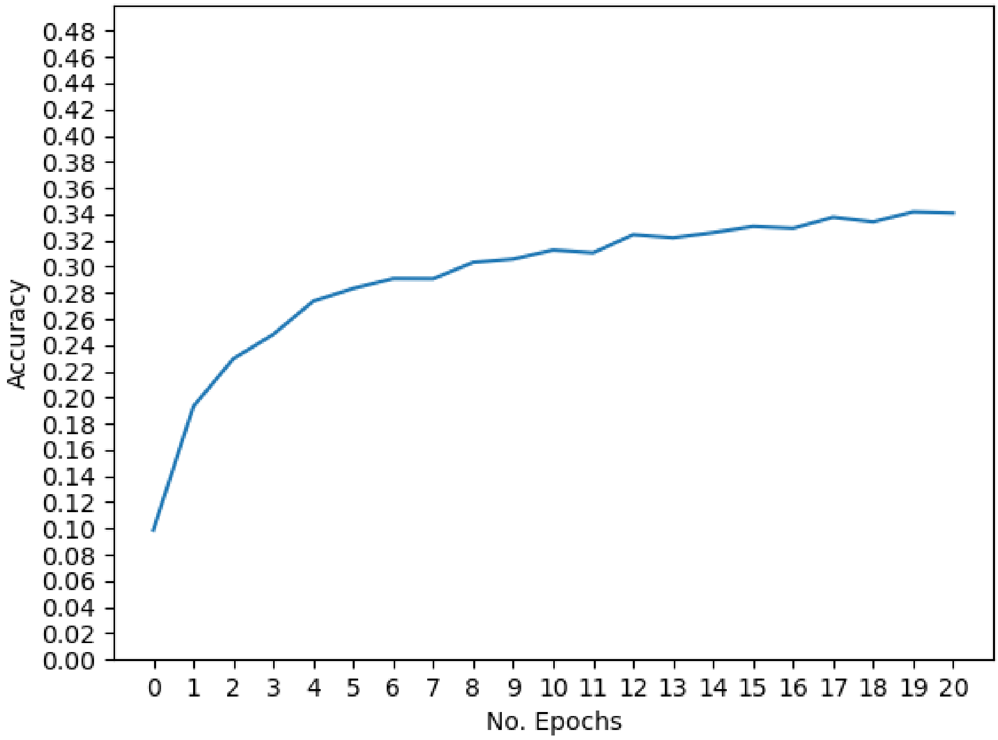

Figure 3: Taking advantage of stuff

Conclusion

Classification is one example of machine learning.

Decision Trees Decision trees are useful for BLAH BLAH BLAH but they are limited in the kind of input they can accept. They are unable to identify visual patterns, so they cannot be used for image classification.

Neural Networks Neural networds are useful for more complex classifcation problems.

Mini-batch size and learning rate both have a significant impact on the performance of a neural network. Too low of a mini-batch size and (EFFECT OF MBS). Too low of a learning rate drastically increases the amount of training time required, while too high of a learning rate tends to get the model stuck in a local optima.

When using a 3-layer neural network (3072 - 30 - 10) to classify images from the cifar-10 dataset, a mini-batch size of 5 and a learning rate of 1 appear to provide the greatest error reduction in the least amount of time.