

KATHOLIEKE UNIVERSITEIT LEUVEN

Prof. Dr. ir. Johan A. K. Suykens

Artificial Neural Networks & Deep Learning

Exercise Reports

Marco Bischoff (R1012984)

May 23, 2024

<i>CONTENTS</i>	2
-----------------	---

Contents

1 Supervised Learning and Generalization	3
1.3 In the Wide Jungle of the Training	3
A Benutzerdokumentation	5
B Introduction	5

1 Supervised Learning and Generalization

1.3 In the Wide Jungle of the Training

Task 1.3.1

What is the impact of the noise (parameter `noise` in the notebook) with respect to the optimization process?

The noise parameter controls the deviation of the training data from the true function. For $noise = 0$, the data exactly matches the true function and the model will converge to the true function quickly. For $noise > 0$, the data is perturbed by noise and the model will take longer to converge. For $noise = 1$, the data is completely random and the model will only converge to the mean of the training data without being able to capture the underlying function.

Task 1.3.2

How does (vanilla) gradient descent compare with respect to its stochastic and accelerated versions?

Vanilla gradient descent is the slowest of the three methods. It computes the gradient of the loss function for the entire training set at each iteration. Stochastic gradient descent is faster than vanilla gradient descent, because it computes the gradient for a random subset of the training data at each iteration. However it has a higher variance in the loss function. Accelerated gradient descent is the fastest of the three methods. It uses a momentum term to speed up convergence and reduce oscillations in the loss function.

Task 1.3.3

How does the size of the network impact the choice of the optimizer?

For small networks, vanilla gradient descent is sufficient, because the computation of the gradient is not very expensive. For larger networks, stochastic gradient descent is more appropriate due to its lower computational cost. Accelerated gradient descent is the best choice for very large networks, because it converges faster than the other two methods.

Task 1.3.4

Discuss the difference between epochs and time to assess the speed of the algorithms. What can it mean to converge fast?

An epoch is a single pass through the entire training set. The time to train a model is the total time it takes to train the model. The number of epochs is a measure of the number of times the model

[1]

References

- [1] Adam Ries. *Rechnung auff der Linihen und Federn*. Hans Schönsperger, Annaberg, 1522.

A Benutzerdokumentation

B Introduction