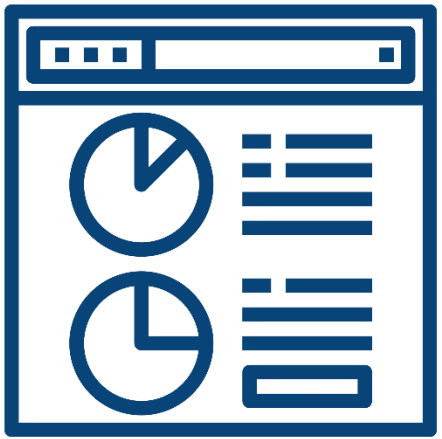
Assignment 2
Introduction

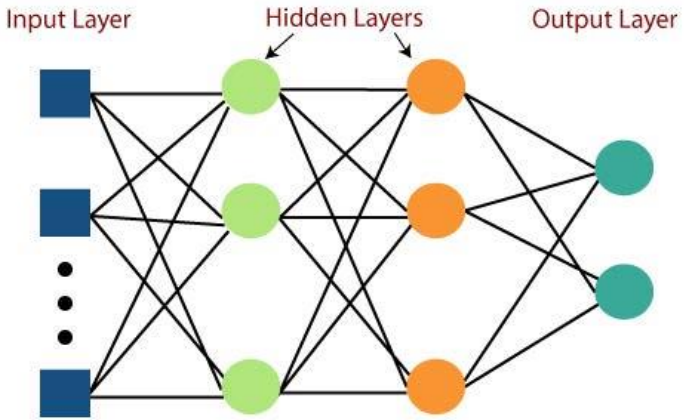
Sebastian Dörrich (MSc)
Deep Learning Exercise
Chair of Explainable Machine Learning (xAI)
28 November 2022

Overview



Overview

1. Softmax Regression
2. 2-Layer Neural Network
3. Activation functions
   a) Sigmoid
   b) ReLU
4. Cross-entropy loss
5. Optimizer
   a) SGD with regularization

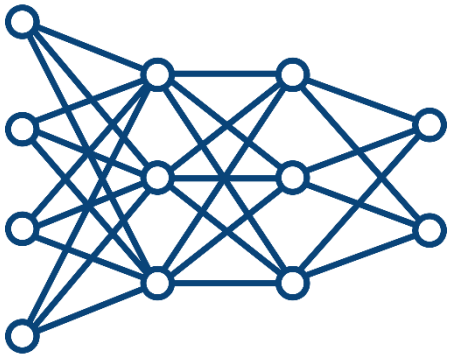


Experiment Report

Experiment Report

- Include all generated plots within the PDF report
- Use the provided template to collect your experiment results
- Add additional pages to the template if necessary

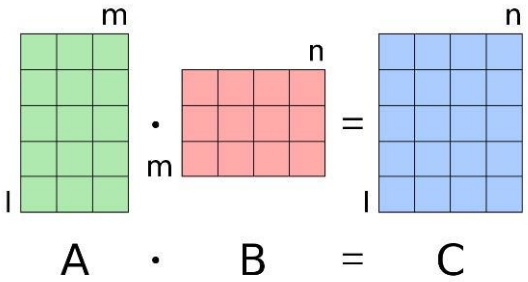
Softmax Regression (SR)



SR – Model Implementation

`X: (N, 784)`

`y: (N, )` ⟶ each is 1 of 10 possible labels



Forward

```
Z = XW
R = ReLU(Z)
S = softmax(R)
L = CE(S, y)
```

Dimensions

`X: (N, 784 ), W: (784, 10)`

`R: (N, 10)` (called “Logits”)

`S: (N, 10)`

L is the loss (scalar)

SR – Model Implementation

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial R}\frac{\partial L}{\partial Z}\frac{\partial Z}{\partial W}$$

Backward

`L = CE(softmax(R), y)`

$$\frac{\partial L}{\partial R} = \frac{\partial CE(softmax(R), y)}{\partial R}$$

$$\frac{\partial L}{\partial Z} = \frac{\partial L}{\partial R}\frac{\partial R}{\partial Z}$$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial Z}\frac{\partial Z}{\partial W}$$

Dimensions

`L` is the loss (scalar)

Same as `R`

Same as `Z`

Same as `W`

SR – Model Implementation

Convenient Derivation for: $\frac{\partial L}{\partial R}$

- $\frac{\partial L}{\partial R}$ is the derivative of the cross entropy loss function w.r.t. the logits.
- Read this article for a detailed explanation

SR – Model Implementation

Tips:

- If you are confused by working with batches, try working through an example with a single sample first and generalize to batches afterward
- When performing matrix operations, think about the dimensions of the desired output and how you can arrive at that given the dimensions of the inputs (e.g. if transposes are needed, which matrix comes first)
- For CE Loss, be sure to take the average across the batch, not the sum