4.1

Layer	Neurons	Activation
Input	N_I=? 2	-
Hidden	N_H=12	Softplus
Output	N_O=? 5	Softmax

Table 1: The architecture of the network.

NI: Inputs are 2D points => number of neurons in the input layer = dimension of the input

NO: C= 5 classes => need 5 neurons in the output layer to represent the probabilites of of each class

4.2

$$\theta = \{\underbrace{\mathbf{W}^{(0)}}_{\substack{\text{Dimension} \\ (NH, NI)}}, \overbrace{\mathbf{b}^{(0)}}^{\substack{\text{Dimension} \\ (NH, 1)}}, \underbrace{\mathbf{W}^{(1)}}_{\substack{\text{Dimension} \\ (NO, NH)}}, \overbrace{\mathbf{b}^{(1)}}^{\substack{\text{Dimension} \\ (NO, 1)}}\}.$$

Total parameter: $\underbrace{\text{size}(W^{(0)})}_{NH \times NI}^{12 \times 2} + \text{size}(b^{(0)})^{12} + \text{size}(W^{(1)})^{12 \times 5} + \text{size}(b^{(1)})^{5}$

NH x NI

(Number of inputs * Number of neurons in hidden layer)

Total parameter = $\underline{101}$

4.3

Layer $l = 0$ (Input)

$$z^{(0)} = x \quad (x \text{ is Input})$$

Layer $l = 1$ (Hidden)

$$z^{(1)} = W^{(0)} x + b^{(0)}$$

$$a^{(1)} = \sigma(z^{(1)}) \quad (\sigma \text{ Softplus activation function})$$

Layer $l = 2$ (Output)

$$z^{(2)} = W^{(1)} a^{(1)} + b^{(1)}$$

$$a^{(2)} = \phi(z^{(2)})$$

Softmax activation func

4.4

$$a_i^{(2)} = \frac{e^{z_i^{(2)}}}{\sum\limits_{j=1}^{N_o} e^{z_j^{(2)}}}$$

$$\frac{\partial a^{(2)}}{\partial z^{(2)}} = \begin{pmatrix} \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} & \frac{\partial a_1^{(2)}}{\partial z_2^{(2)}} & \cdots & \frac{\partial a_1^{(2)}}{\partial z_{N_o}^{(2)}} \\ \frac{\partial a_2^{(2)}}{\partial z_1^{(2)}} & \ddots & & \\ \vdots & & \ddots & \\ \frac{\partial a_{N_o}^{(2)}}{\partial z_1^{(2)}} & & & \frac{\partial a_{N_o}^{(2)}}{\partial z_{N_o}^{(2)}} \end{pmatrix}$$

for $i=j$:

quotient rule: $h(z_i^{(2)}) = e^{z_i^{(2)}}$ $\quad h'(z_i^{(2)}) = e^{z_i^{(2)}}$

$g(z) = \sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}$ $\quad g'(z) = e^{z_j^{(2)}}$

$$\frac{\partial a_i^{(2)}}{\partial z_j^{(2)}} = \frac{e^{z_i^{(2)}} \sum\limits_{k=1}^{N_o} e^{z_k^{(2)}} - e^{z_i^{(2)}} e^{z_j^{(2)}}}{\left(\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}\right)^2} =$$

$$= \frac{e^{z_i^{(2)}}\left(\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}} - e^{z_j^{(2)}}\right)}{\left(\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}\right)^2} = \underbrace{\frac{e^{z_i^{(2)}}}{\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}}}_{a_i^{(2)}} \cdot \left(\underbrace{\cancel{\frac{\sum_{k=1}^{N_o} e^{z_k^{(2)}}}{\sum_{k=1}^{N_o} e^{z_k^{(2)}}}}}_{1} - \underbrace{\frac{e^{z_j^{(2)}}}{\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}}}_{a_j^{(2)}} \right)$$

$$= a_i^{(2)}\left(1 - a_j^{(2)}\right) = a_i^{(2)}\left(1 - a_i^{(2)}\right)$$

for $i \neq j$:

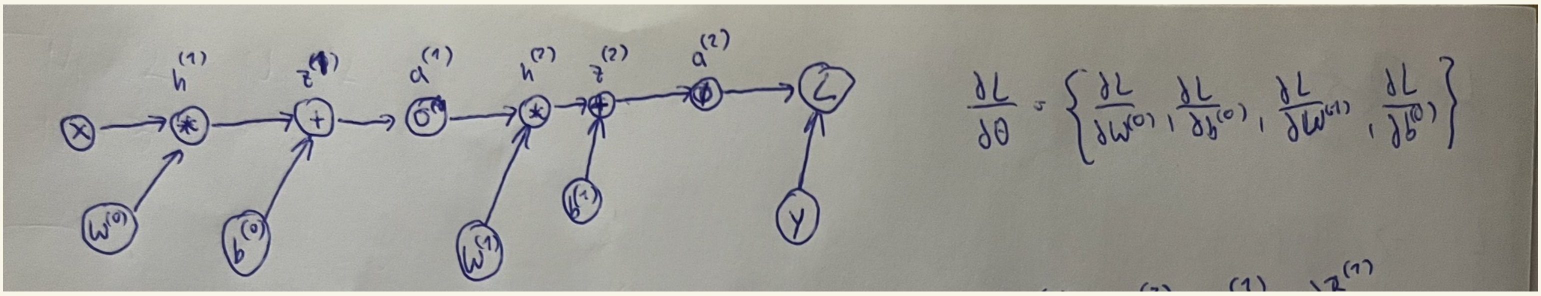
quotient rule: $h(z_i^{(2)}) = e^{z_i^{(2)}}$ $\quad h'(z_i^{(2)}) = 0$

$g(z) = \sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}$ $\quad g'(z) = e^{z_j^{(2)}}$

$$\frac{\partial a_i^{(2)}}{\partial z_j^{(2)}} = \frac{0 - e^{z_i^{(2)}} e^{z_j^{(2)}}}{\left(\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}\right)^2} = -\underbrace{\frac{e^{z_i^{(2)}}}{\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}}}_{a_i^{(2)}} \cdot \underbrace{\frac{e^{z_j^{(2)}}}{\sum\limits_{k=1}^{N_o} e^{z_k^{(2)}}}}_{a_j^{(2)}} = -a_i^{(2)} a_j^{(2)}$$

$$\frac{\partial a_i^{(2)}}{\partial z_j^{(2)}} = \begin{cases} a_i^{(2)}\left(1 - a_i^{(2)}\right) & \text{for } i=j \\ -a_i^{(2)} a_j^{(2)} & \text{for } i \neq j \end{cases}$$

4.5



$$L(\theta) = \frac{1}{S} \sum_{s=1}^{S} \ell(a^{(2)s}, y^s)$$

$$h^{(1)} = W^{(0)} x + b^{(0)}$$

$$a^{(1)} = \sigma(h^{(1)})$$

$$h^{(2)} = W^{(1)} a^{(1)} + b^{(1)}$$

$$a^{(2)} = \phi(h^{(2)})$$

* $W^{(1)}$

$$\frac{dL}{dW^{(1)}} = \frac{1}{S} \sum_{s=1}^{S} \frac{d\ell(a^{(2)s}, y^s)}{dW^{(1)}}$$

$$= \frac{1}{S} \sum_{s=1}^{S} a^{(1)s} (\phi(h^{(2)s} - y^s)$$

Dimension : $NO \times NH = 5 \times 12$

* $b^{(1)}$

$$\frac{dL}{db^{(1)}} = \frac{1}{S} \sum_{s=1}^{S} \frac{d\ell(a^{(2)s}, y^s)}{db^{(1)}}$$

$$= \frac{1}{S} \sum_{s=1}^{S} (\phi(h^{(2)s} - y^s)$$

Dimension : $NO \times 1 = 5 \times 1$

* $W^{(0)}$

$$\frac{dL}{dW^{(0)}} = \frac{1}{S}\sum_{s=1}^{S} \frac{d\ell(a^{(2)s}, y^s)}{dW^{(0)}}$$

$$= \frac{1}{S}\sum_{s=1}^{S} \sigma'(h^{(1)s})\, W^{(1)T} \cdot (\phi(h^{(2)s}) - y^s) \cdot x^{sT}$$

Dimension : $NH \times NI = 12 \times 2$

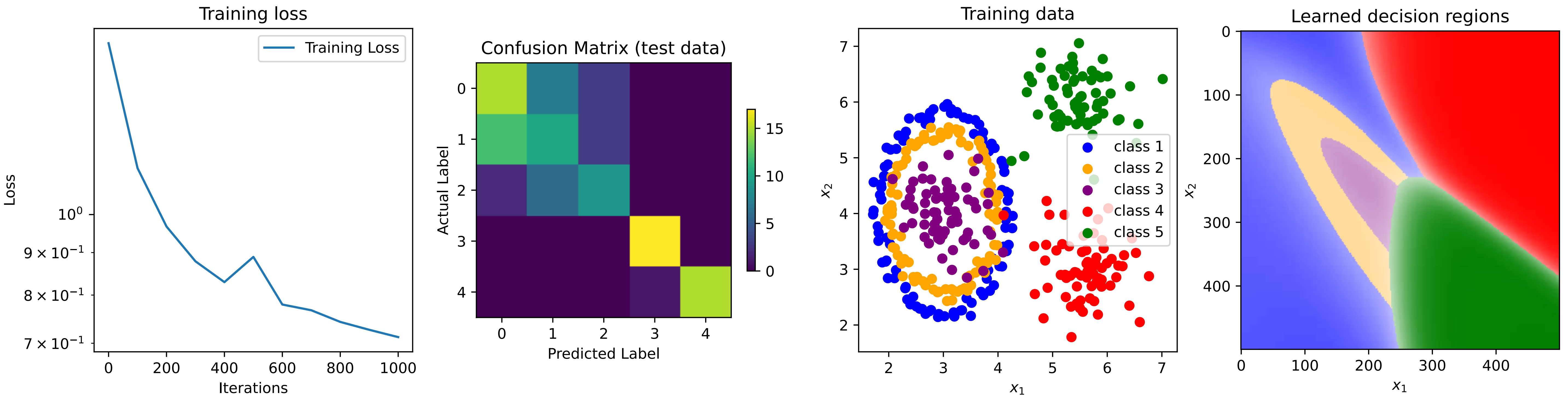
* $b^{(0)}$

$$\frac{dL}{db^{(0)}} = \frac{1}{S}\sum_{s=1}^{S} \frac{d\ell(a^{(2)s}, y^s)}{db^{(0)}}$$

$$= \frac{1}{S}\sum_{s=1}^{S} \sigma'(h^{(1)s})\, W^{(1)T} \cdot (\phi(h^{(2)s}) - y^s)$$

Dimension : $NH \times 1 = 12 \times 1$

6.1



Training loss
Training Loss
Loss
10^0
9×10^{-1}
8×10^{-1}
7×10^{-1}
0
200
400
600
800
1000
Iterations
Confusion Matrix (test data)
Actual Label
Predicted Label
0
1
2
3
4
15
10
5
0
Training data
class 1
class 2
class 3
class 4
class 5
x_1
x_2
Learned decision regions
0
100
200
300
400
x_1
x_2

Index der Kommentare

1.1 -0.5: test results not reported

1.2 -2: nothing reported

6.1 1.8: -1