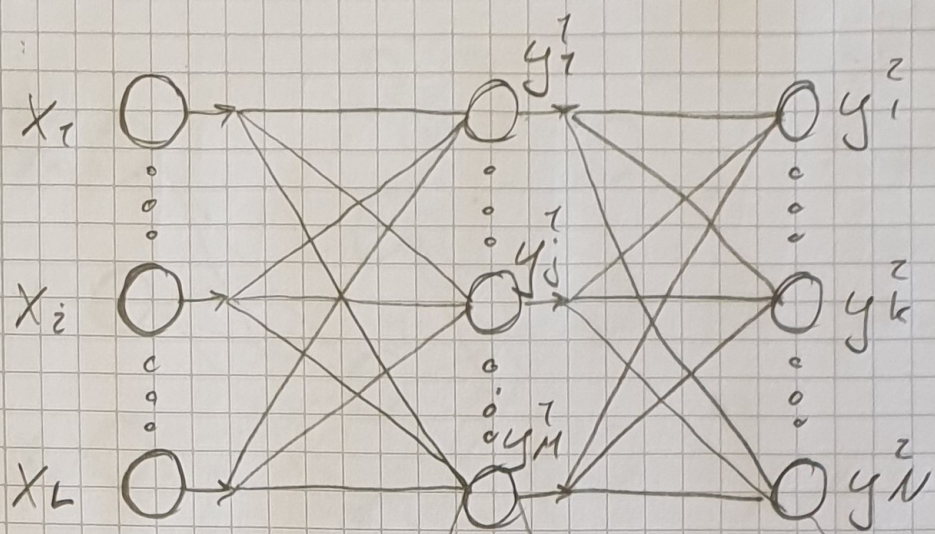
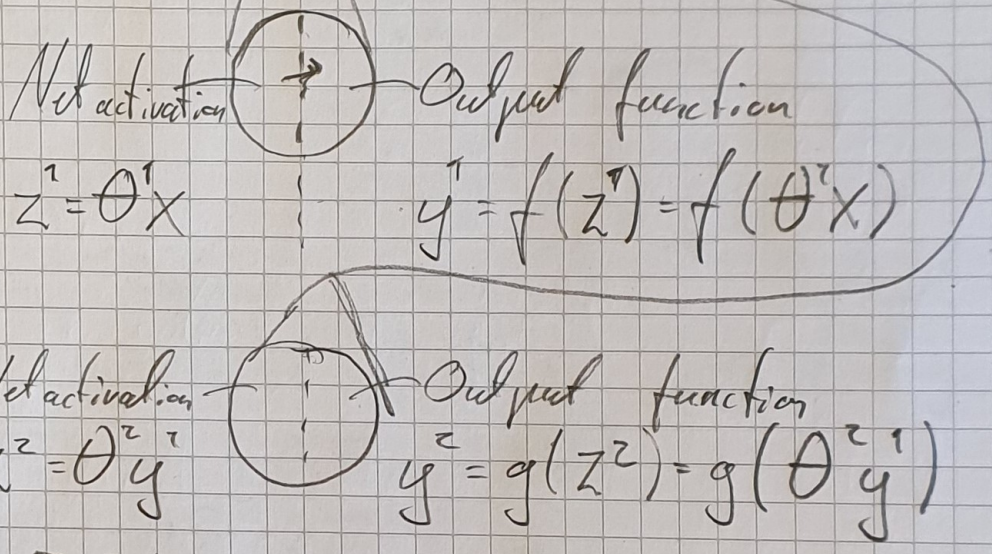
ELE520 Exercise 4

P7.





$$y^2 = g[\theta^2 f(\theta^1 x)]$$

If both g and f are linear we have:

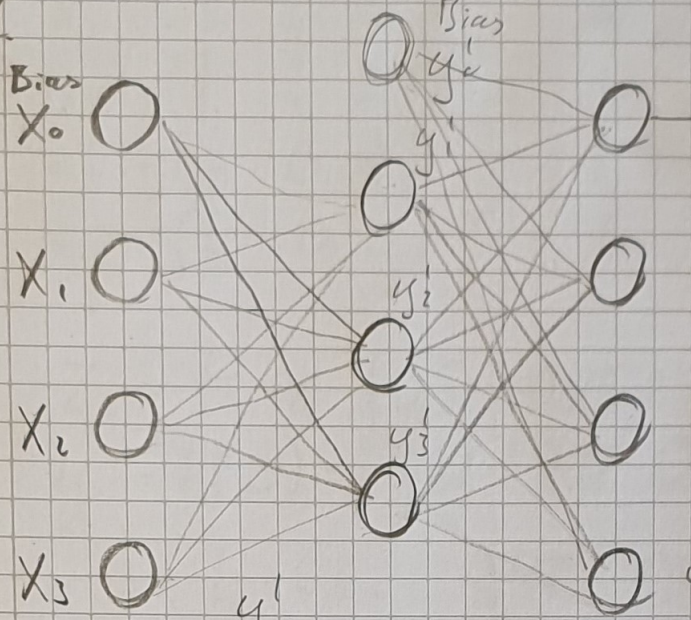
$$y^2 = G\theta^2 F\theta^1 x$$

Weight matrices and output scaling can be combined into a single matrix $\theta^3 = G\theta^2 F\theta^1$

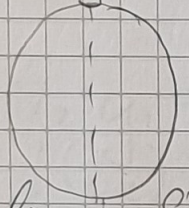
$$y^2 = \theta^3 x$$

ELE520 Exercise 4

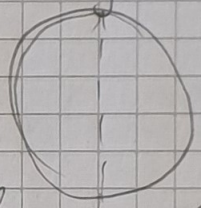
Problem 2

a) 

(Integrating biases into main structure)



$y_1 = y_0^1 + y_1$



activation $z_1 = \theta^1 x$ | output $y_1 = f(\theta^1 x)$ | Activation $z_2 = \theta^2 y_1$ | Output $y_2 = f(\theta^2 y_1)$

$$\theta^1 = \begin{bmatrix} 0.5 & 0 & -0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0 \\ 0.5 & -0.5 & 0 & 0.5 \end{bmatrix}$$

$$\theta^2 = \begin{bmatrix} -0.5 & 0.5 & -0.5 & 0.5 \\ 0.5 & 0 & -0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & -0.5 \end{bmatrix}$$

ELE520 Exercise 4 Vai 4/8.

P2 b) $x_1 = (1 \ \frac{1}{4} \ \frac{1}{4})^T$ A

$x_2 = (1 \ \frac{1}{4} \ 0)^T$ P

$x_3 = (\frac{1}{2} \ 0 \ \frac{1}{4})^T$ C

$x_4 = (\frac{1}{2} \ 0 \ 0)^T$ F

c) $z_1 = \theta_1 x = \begin{bmatrix} 0.5 & -0.125 & 0.125 \end{bmatrix} (z) = \frac{1}{1+e^{-z}}$

$y_1 = f(z_1) = (0.62 \ 0.47 \ 0.53)$

$y_1 = (1 \ 0.62 \ 0.47 \ 0.53)$

$z_2 = \theta_2 y_1 = (-0.16 \ 0.53 \ -0.57 \ 0.54)$

$y_2 = f(z_2) = (0.46 \ 0.63 \ 0.36 \ 0.63)$

$J(\theta) = \frac{1}{2} \| y - y^2 \|^2 \quad , \quad y = (1 \ 0 \ 0 \ 0)$

$J(\theta) = 1.22$

d) $f'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$

$\delta^2 = diag(f'(z^2))(y - y^2) = (0.13 \ -0.15 \ -0.08 \ -0.15)$

$\delta^1 = diag(f'(z^1))(\theta^{2T} \delta^2) = (0 \ 0.008 \ -0.009 \ 0.017)$

$\Delta\theta^2 = \mu \delta^2 y^{1T} = \begin{bmatrix} 0.13 & \cdots & 0.07 \\ \vdots & & \vdots \\ -0.15 & \cdots & -0.08 \end{bmatrix}_{4 \times 4}$

$\Delta\theta^1 = \mu \delta^1 x^T = \begin{bmatrix} 0.008 & \cdots & 0.002 \\ \vdots & & \\ 0.017 & \cdots & 0.004 \end{bmatrix}_{4 \times 3}$

ELE520 Exercise 4

$$\Theta^{2}_{new} = \begin{bmatrix} -0.36 & 0.58 & -0.44 & 0.57 \\ 0.35 & -0.09 & -0.56 & 0.42 \\ -0.58 & -0.55 & 0.46 & -0.04 \\ 0.35 & 0.41 & -0.07 & -0.58 \end{bmatrix}$$

$$\Theta^{1}_{new} = \begin{bmatrix} 0.51 & 0.008 & -0.498 & 0.502 \\ 0.504 & 0.491 & -0.502 & -0.002 \\ 0.517 & -0.483 & 0.004 & 0.504 \end{bmatrix}$$

```
import numpy as np
```

```
theta1 = np.asmatrix([[.5,0,-.5,.5],[-.5,.5,-.5,0],[.5,-.5,0,.5]])
theta2 = np.asmatrix([[-.5,.5,-.5,.5],[.5,0,-.5,.5],[-.5,-.5,.5,0],[.5,.5,0,-.5]])
x = np.asmatrix([1,1,0.25,0.25])
x
```

```
matrix([[1.  , 1.  , 0.25, 0.25]])
```

```
z1 = theta1*x.T
z1 = np.r_[[[np.inf]], z1]
z1
```

```
matrix([[   inf],
        [ 0.5  ],
        [-0.125],
        [ 0.125]])
```

```
def f(z):
    return 1/(1+np.exp(-z))
```

```
y1 = f(z1)
y1
```

```
matrix([[1.        ],
        [0.62245933],
        [0.46879063],
        [0.53120937]])
```

```
z2 = theta2 * y1
z2
```

```
matrix([[-0.15756096],
        [ 0.53120937],
        [-0.57683435],
        [ 0.54562498]])
```

```
y2 = f(z2)
y2
```

```
matrix([[0.46069105],
        [0.62976513],
        [0.35966133],
        [0.63311996]])
```

```
y = np.asmatrix([1,0,0,0]).T

J = 0.5* np.sum(np.square(y - y2))
J
```

```
def f_diff(z):
    return np.exp(-z)/np.square((1+np.exp(-z)))
```

+ Code + Markdown

```
np.asarray(z2).flatten()
```

```
array([-0.15756096,  0.53120937, -0.57683435,  0.54562498])
```

```
diag2 = np.diag(f_diff(np.asarray(z2).flatten()))
diag2
```

```
array([[0.24845481, 0.        , 0.        , 0.        ],
       [0.        , 0.23316101, 0.        , 0.        ],
       [0.        , 0.        , 0.23030506, 0.        ],
       [0.        , 0.        , 0.        , 0.23227908]])
```

```
delta2 = diag2 * (y-y2)
delta2
```

```
matrix([[ 0.1339939 ],
        [-0.14683667],
        [-0.08283182],
        [-0.14706052]])
```

```
diag1 = np.diag(f_diff(np.asarray(z1).flatten()))
diag1
```

```
array([[0.        , 0.        , 0.        , 0.        ],
       [0.        , 0.23500371, 0.        , 0.        ],
       [0.        , 0.        , 0.24902598, 0.        ],
       [0.        , 0.        , 0.        , 0.24902598]])
```

+ Code + Markdown

```
delta1 = diag1 * theta2.T * delta2
delta1 = delta1[1:]
delta1
```

```
matrix([[ 0.00819754],
        [-0.00871455],
        [ 0.01671185]])
```

```
mu = 1
```

```
d_theta2 = mu * delta2 * y1.T
d_theta2
```

```
matrix([[ 0.1339939 ,  0.08340575,  0.06281508,  0.07117882],
        [-0.14683667, -0.09139986, -0.06883566, -0.07800102],
        [-0.08283182, -0.05155944, -0.03883078, -0.04400104],
        [-0.14706052, -0.09153919, -0.06894059, -0.07811993]])
```

```
    d_theta1 = mu * delta1 * x
    d_theta1
```

```
matrix([[ 0.00819754,  0.00819754,  0.00204939,  0.00204939],
        [-0.00871455, -0.00871455, -0.00217864, -0.00217864],
        [ 0.01671185,  0.01671185,  0.00417796,  0.00417796]])
```

```
    theta2_new = theta2 + d_theta2
    theta2_new
```

```
matrix([[-0.3660061 ,  0.58340575, -0.43718492,  0.57117882],
        [ 0.35316333, -0.09139986, -0.56883566,  0.42199898],
        [-0.58283182, -0.55155944,  0.46116922, -0.04400104],
        [ 0.35293948,  0.40846081, -0.06894059, -0.57811993]])
```

```
    theta1_new = theta1 + d_theta1
    theta1_new
```

```
matrix([[ 0.50819754,  0.00819754, -0.49795061,  0.50204939],
        [-0.50871455,  0.49128545, -0.50217864, -0.00217864],
        [ 0.51671185, -0.48328815,  0.00417796,  0.50417796]])
```