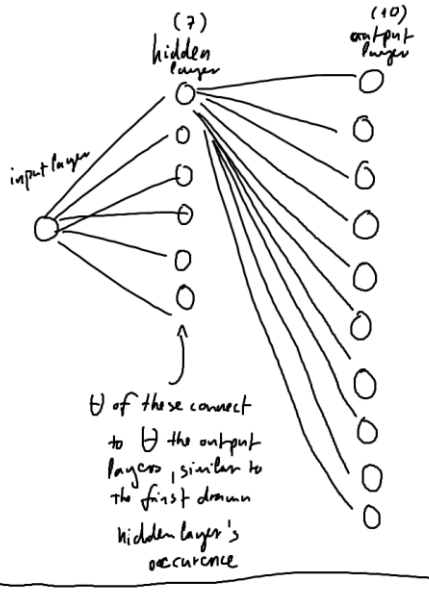
# Computer Vision Exercise 8

(1)




onward number $^2$ is number (2)
number $^1$ is number (1)

I. $m=1$ $\partial x$? and $\partial t$?

$$\Rightarrow E = \frac{1}{m}\sum_{j=1}^{m} -t_j \cdot \log(y_j) \quad \rightarrow E = -t \cdot \log(y) = -t \log(\sigma(Wx)) = -t \log \frac{1}{1+e^{-Wx}}$$

$$\frac{\partial E}{\partial t} = -\log \frac{1}{1+e^{-Wx}} \qquad \frac{\partial E}{\partial x} = -\frac{t}{\sigma(Wx)} \cdot \frac{\partial \sigma(Wx)}{\partial x} = -t(1+e^{-Wx}) \cdot W \frac{1}{1+e^{-Wx}} \left(1 - \frac{1}{1+e^{-Wx}}\right)$$

$$= -tW \frac{e^{-Wx}}{1+e^{-Wx}}$$

II $\frac{\partial E}{\partial z_i^2} = (y^{(2)} - t)^T$ $\qquad s(z_i^2) = \frac{e^{z_i^2}}{\sum_{a=1}^{n_2} e^{z_a^2}} = y_i^2$ $\qquad \sum t_j = 1$

chain rule:

$$\frac{\partial E}{\partial z_i^2} = \sum_{j=1}^{n} \frac{\partial E_j}{\partial y_j^2} \cdot \frac{\partial y_j^2}{\partial z_i^2}$$

$$= \frac{\partial(-t_i \cdot \log y_i)}{\partial z_i^2} = -\frac{t_i}{y_i}$$

if $i=j \Rightarrow y_i^2(1-y_i^2)$

if $i \neq j \Rightarrow -y_i^2 y_j^2$

$$\Leftrightarrow \sum_{i \neq j}^{n} \frac{\partial E_i}{\partial y_j^2} \frac{\partial y_j^2}{\partial z_i^2} + \sum_{i=j}^{n} \frac{\partial E_i}{\partial y_j^2} \frac{\partial y_j^2}{\partial z_i^2} = \sum_{i=j}^{n} \frac{t_i}{y_j^2} \cdot (-y_i^2 y_j^2) + \sum_{i \neq j}^{n} \frac{t_i}{y_j^2} y_i^2 (1-y_i^2)$$

$$= \sum_{i=j}^{n} \frac{t_i}{y_i^2} \cdot (y_i^2 y_j^2) - \frac{t_i}{y_i^2} y_i^2 (1-y_i^2) = y_i - t_j$$

III. $\frac{\partial E}{\partial z^1} = (y^2 - t)^T W^2$ $\qquad \frac{\partial E}{\partial z^2} = (y^2 - t)^T$

$$\frac{\partial E}{\partial y^1} = \frac{\partial E}{\partial z^2} \cdot \frac{\partial z^2}{\partial y^1} = (y^2 - t)^T \cdot W^2$$

$\frac{\partial z_i^2}{\partial y_j^1} = W_{ij}^2$

IV. $\frac{\partial E}{\partial w_{uv}^2} = (y_u^2 - t_u) y_v^1$ $\qquad \frac{\partial E}{\partial z_u^2} = y_u^2 - t_u$

$\frac{\partial z_u^2}{\partial w_{uv}^2} = y_v^1$ $\qquad \frac{\partial E}{\partial w_{uv}^2} = \frac{\partial E}{\partial z_u^2} \cdot \frac{\partial z_u^2}{\partial w_{uv}^2} \Rightarrow (y_u^2 - t_u) y_v^1$

V. $\frac{\partial y^1}{\partial z^1} = \mathrm{diag}(y^1 * (1 - y^1))$

log func: $y = \sigma(z) = \frac{1}{1+e^{-z}}$

$$\frac{\partial \sigma(z)}{\partial z} = \frac{-(-1)e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \frac{e^{-z}}{(1+e^{-z})^2} = \sigma(z)(1-\sigma(z))$$

$y_i^1 = \sigma(z^1) \Rightarrow \frac{\partial y_i^1}{\partial z_i^1} = y_i^1 (1 - y_i^1) \Rightarrow \frac{\partial y^1}{\partial z^1} = \mathrm{diag}(y^1 * (1 - y^1))$

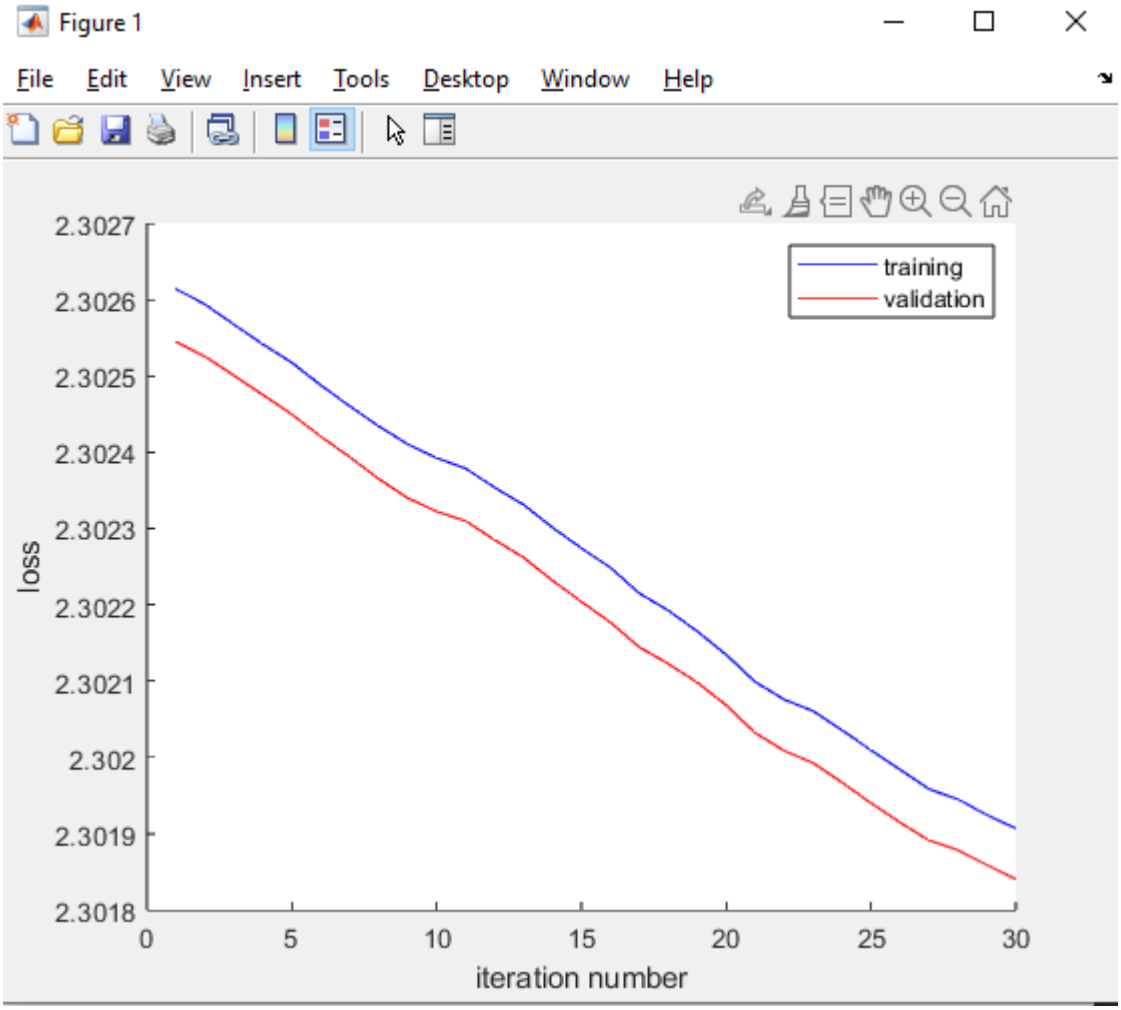
VI. $\frac{\partial E}{\partial z^1} = (y^2 - t)^T W^2 \mathrm{diag}(y^1 * (1 - y^1))$

$= \frac{\partial E}{\partial y^2} \cdot \frac{\partial y^2}{\partial z^1} =$

$(y^2 - t) \cdot (W^2 \cdot \frac{\partial y^1}{\partial z^1})$

VII. $\frac{\partial E}{\partial w_{uv}^1} = \frac{\partial E}{\partial z_u^1} x_v$ $\qquad \frac{\partial E}{\partial W^1} = \left(\frac{\partial E}{\partial z^1}\right)^T x^T$

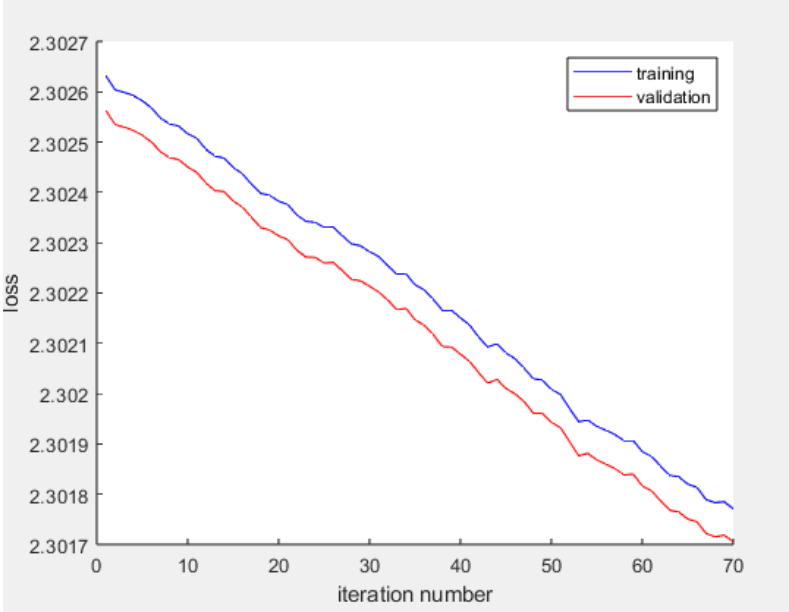
1. chain rule

2. 

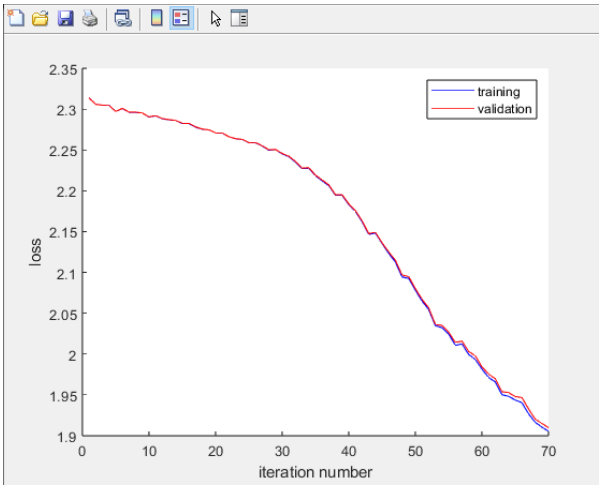

```
The total loss on the training data is 2.301907
The classification loss (i.e. without weight decay) on the training data is 2.301907
The classification error rate on the training data is 0.889000

The total loss on the validation data is 2.301841
The classification loss (i.e. without weight decay) on the validation data is 2.301841
The classification error rate on the validation data is 0.895000

The total loss on the test data is 2.301865
The classification loss (i.e. without weight decay) on the test data is 2.301865
The classification error rate on the test data is 0.887333
```

3. .

```
The total loss on the training data is 2.301771
The classification loss (i.e. without weight decay) on the training data is 2.301771
The classification error rate on the training data is 0.888000

The total loss on the validation data is 2.301706
The classification loss (i.e. without weight decay) on the validation data is 2.301706
The classification error rate on the validation data is 0.893000

The total loss on the test data is 2.301728
The classification loss (i.e. without weight decay) on the test data is 2.301728
The classification error rate on the test data is 0.886667
```




```
>> a2(0, 10, 70, 0.5, 0, false, 4)
Now testing the gradient on the whole training set... Gradient test passed for hid_to_class. Gradient test
After 7 optimization iterations, training data loss is 2.295996, and validation data loss is 2.296220
After 14 optimization iterations, training data loss is 2.285828, and validation data loss is 2.286007
After 21 optimization iterations, training data loss is 2.270682, and validation data loss is 2.270843
After 28 optimization iterations, training data loss is 2.249667, and validation data loss is 2.250156
After 35 optimization iterations, training data loss is 2.218406, and validation data loss is 2.219122
After 42 optimization iterations, training data loss is 2.162092, and validation data loss is 2.163039
After 49 optimization iterations, training data loss is 2.092659, and validation data loss is 2.094904
After 56 optimization iterations, training data loss is 2.010721, and validation data loss is 2.014252
After 63 optimization iterations, training data loss is 1.950202, and validation data loss is 1.953984
After 70 optimization iterations, training data loss is 1.905315, and validation data loss is 1.909771
Now testing the gradient on just a mini-batch instead of the whole training set... Gradient test passed f

The total loss on the training data is 1.905315
The classification loss (i.e. without weight decay) on the training data is 1.905315
The classification error rate on the training data is 0.792000

The total loss on the validation data is 1.909771
The classification loss (i.e. without weight decay) on the validation data is 1.909771
The classification error rate on the validation data is 0.791000

The total loss on the test data is 1.900503
The classification loss (i.e. without weight decay) on the test data is 1.900503
The classification error rate on the test data is 0.785889
```




[...]