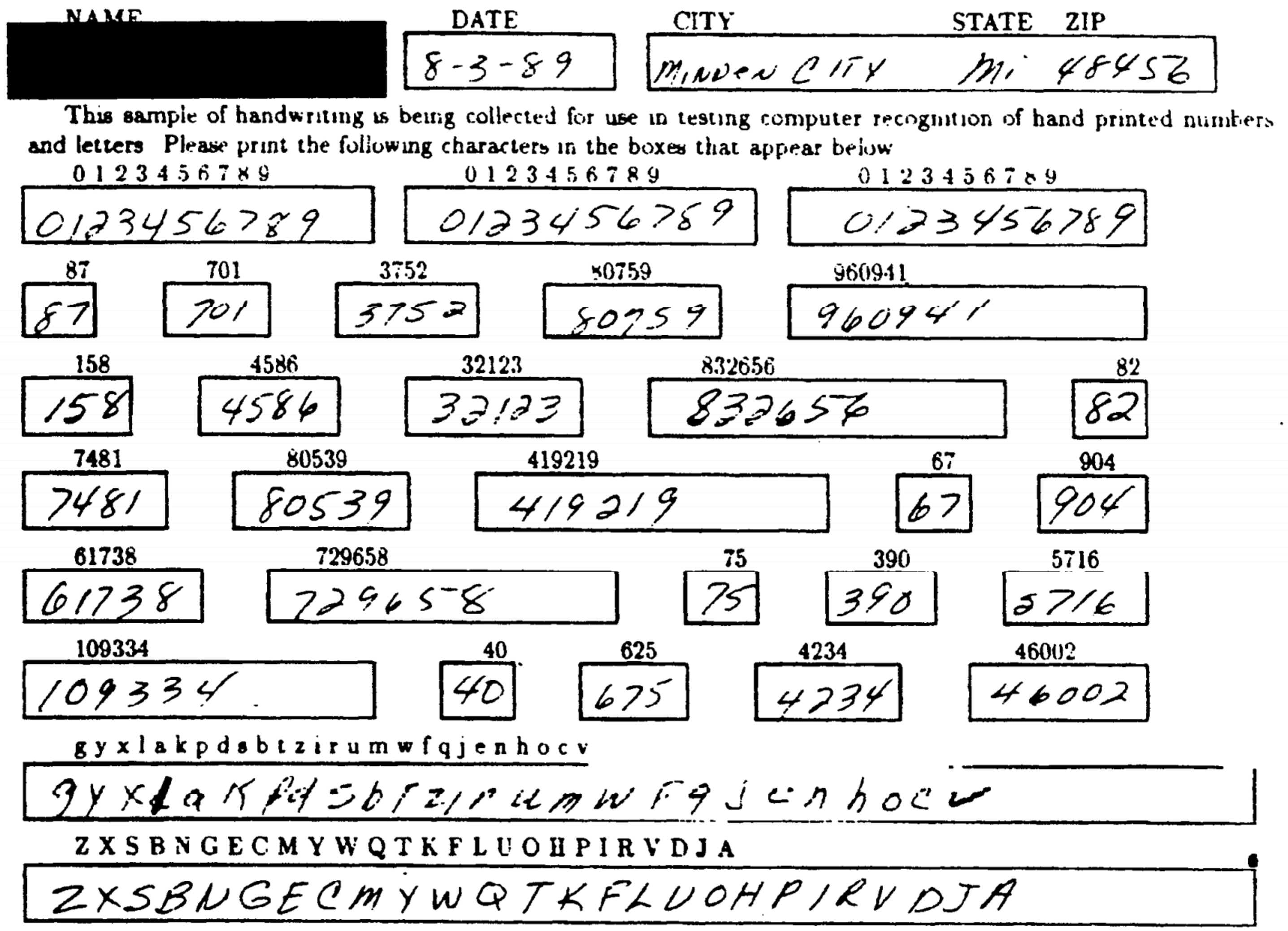
MLP for MNIST

by 신경식

MNIST Dataset



NAME [redacted] DATE 8-3-89 CITY MINDEN CITY STATE MI ZIP 48456

This sample of handwriting is being collected for use in testing computer recognition of hand printed numbers and letters. Please print the following characters in the boxes that appear below.

0 1 2 3 4 5 6 7 8 9 | 0 1 2 3 4 5 6 7 8 9 | 0 1 2 3 4 5 6 7 8 9

87 | 701 | 3752 | 80759 | 960941

158 | 4586 | 32123 | 832656 | 82

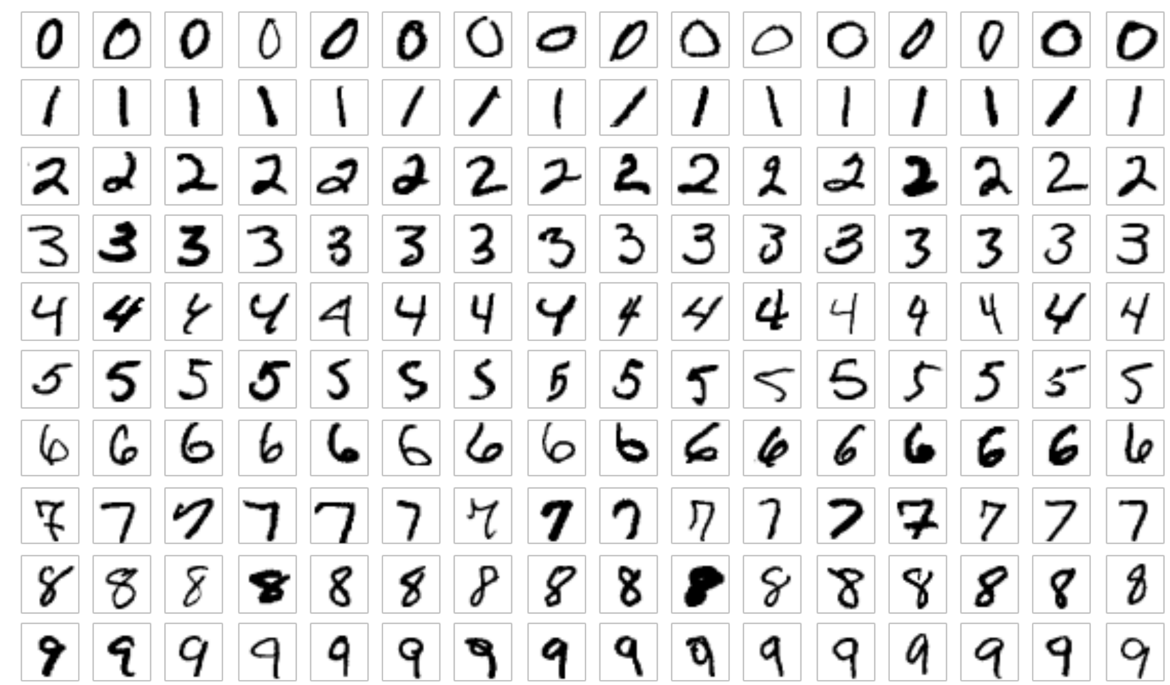
7481 | 80539 | 419219 | 67 | 904

61738 | 729658 | 75 | 390 | 5716

109334 | 40 | 625 | 4234 | 46002

g y x l a k p d s b t z i r u m w f q j e n h o c v

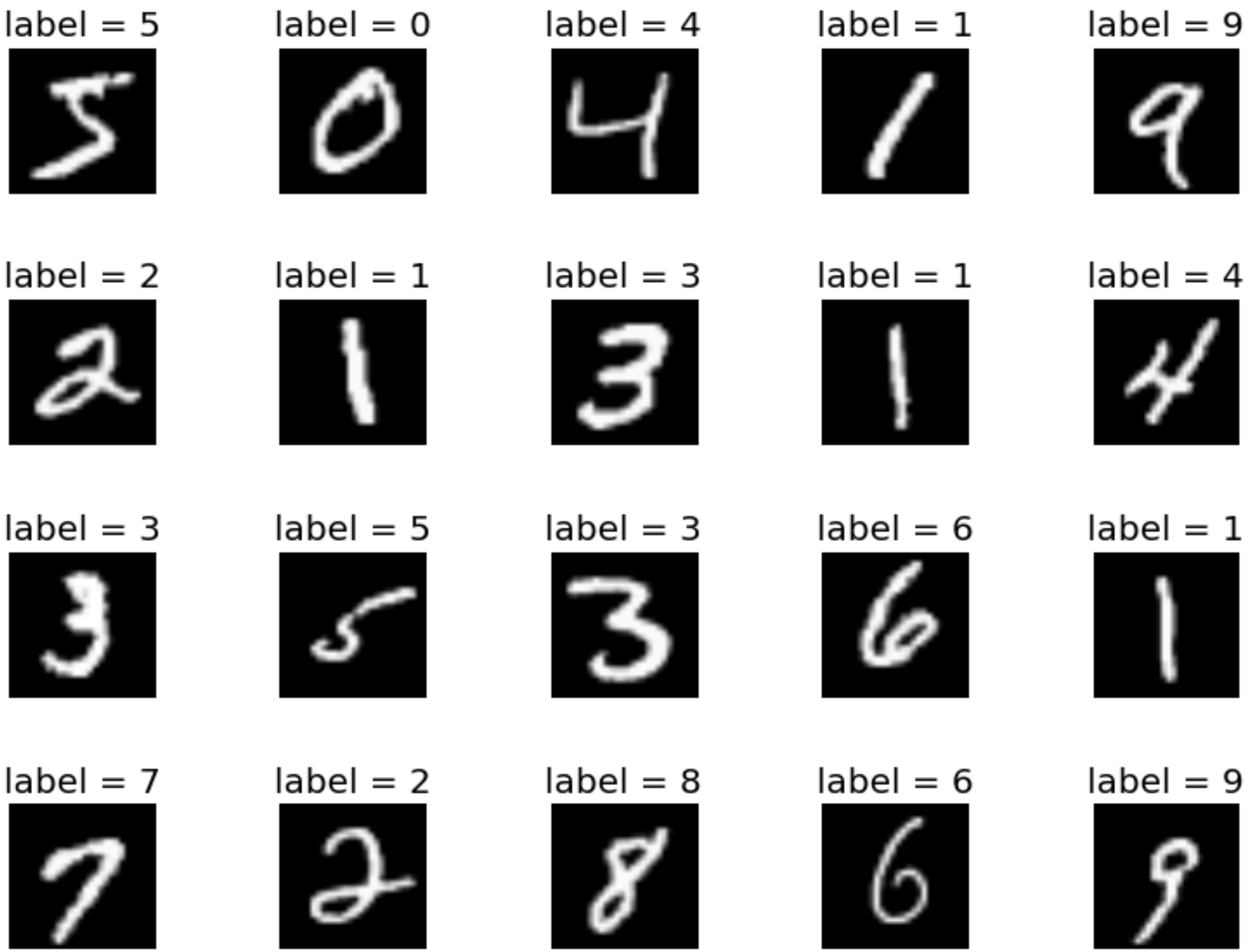
Z X S B N G E C M Y W Q T K F L U O H P I R V D J A



[1] https://www.nist.gov/system/files/documents/srd/nistsd19.pdf

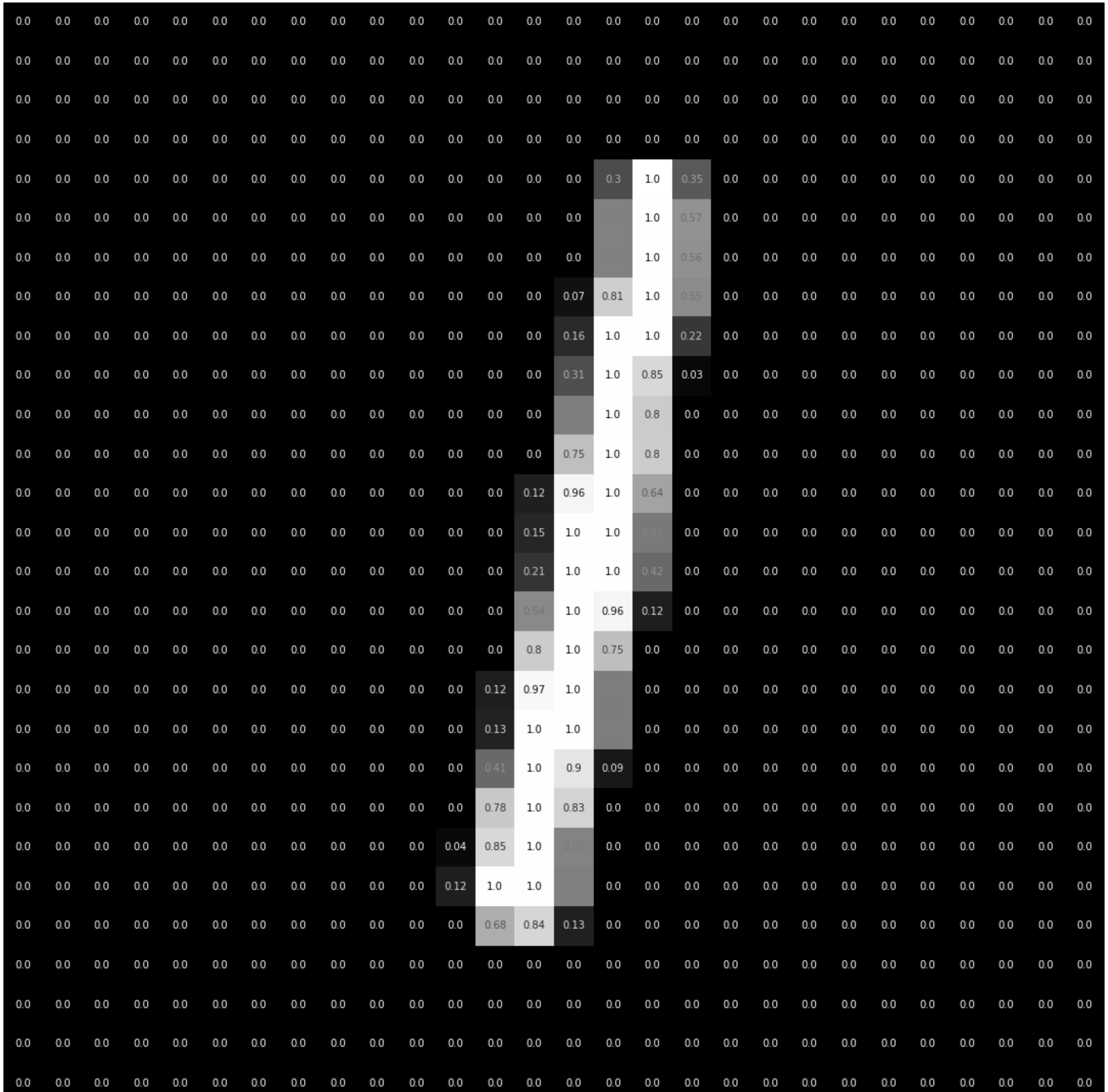
[2] https://ko.wikipedia.org/wiki/MNIST_%EB%8D%B0%EC%9D%B4%ED%84%B0%EB%B2%A0%EC%9D%B4%EC%8A%A4

MNIST Dataset

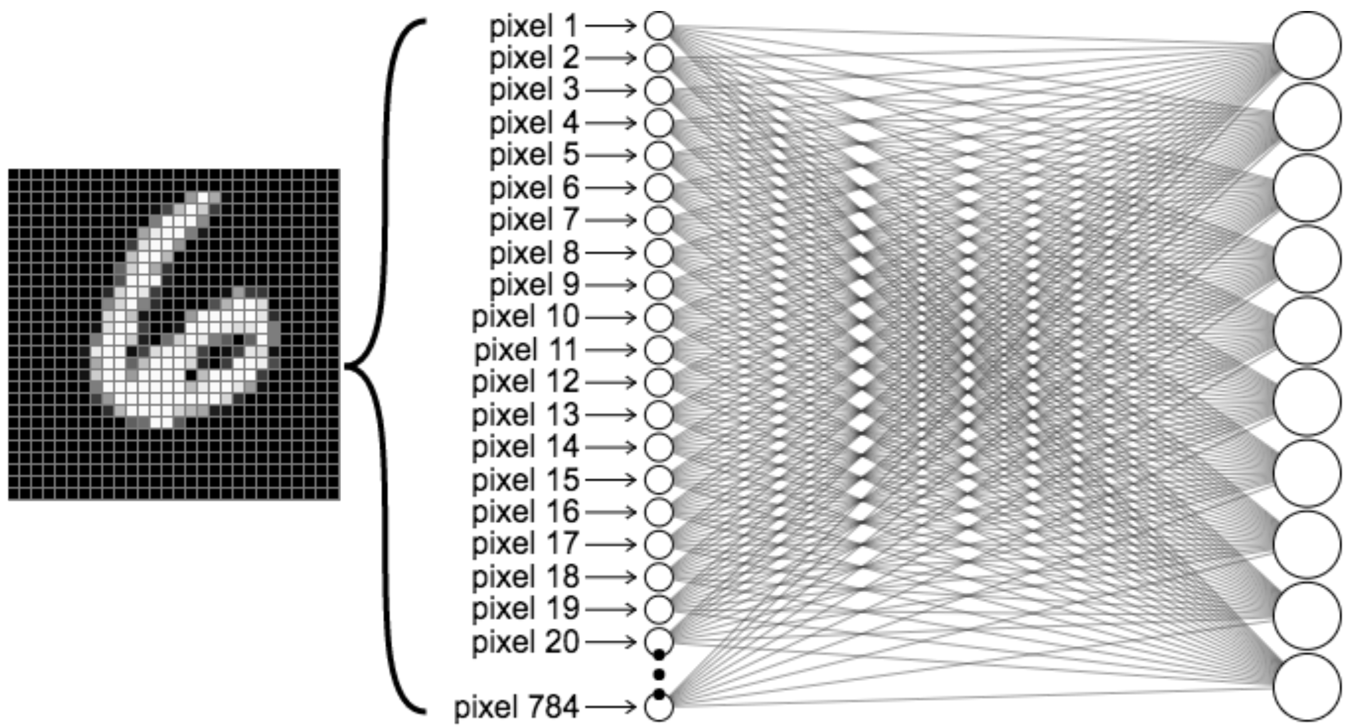


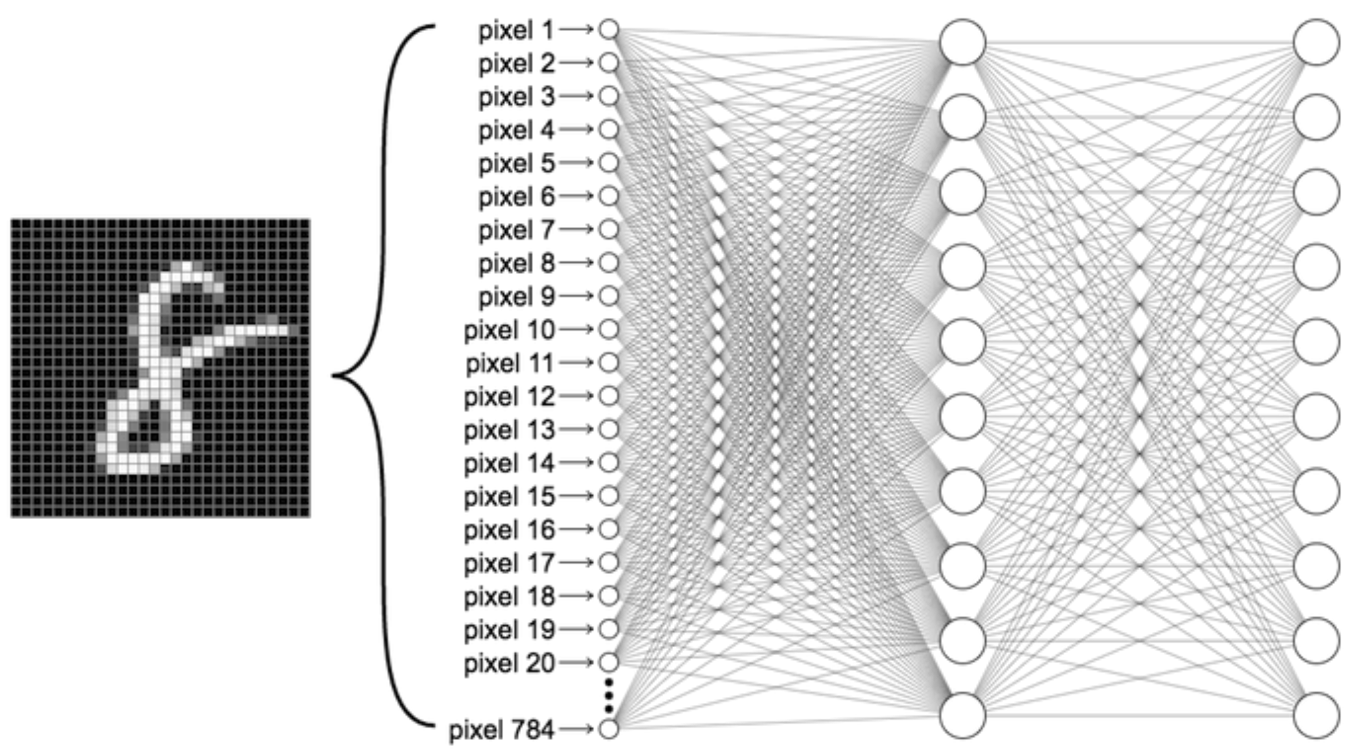
[1] https://corochann.com/mnist-dataset-introduction-1138.html

MNIST Dataset



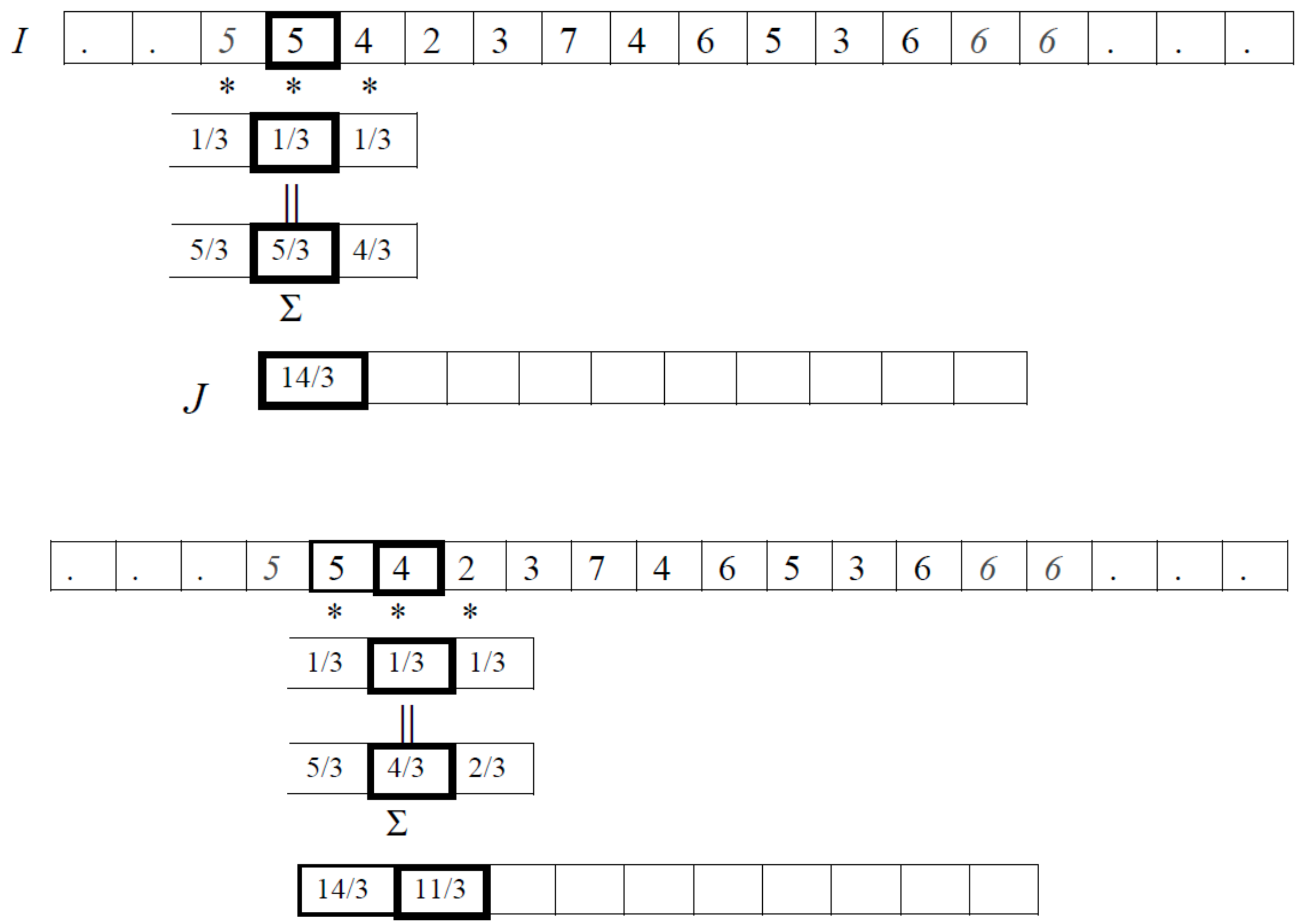
Handwritten Digit Recognition with MLP





[1] https://ml4a.github.io/ml4a/looking_inside_neural_nets/

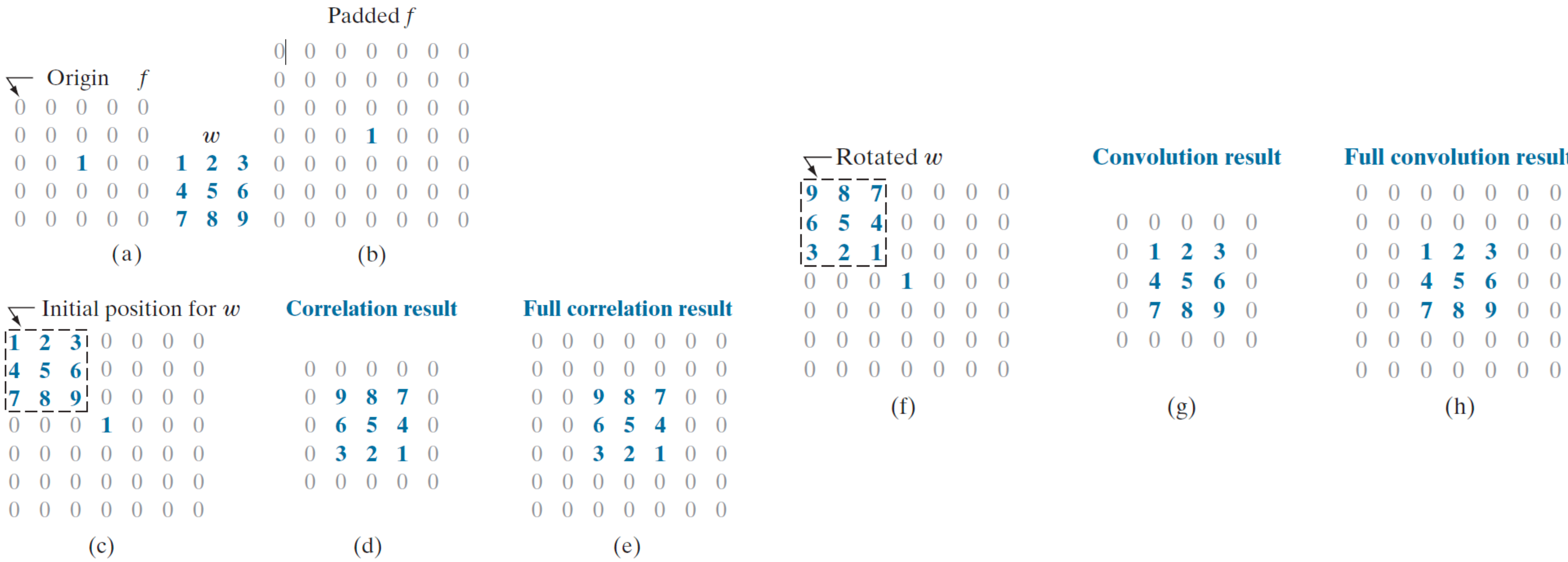
Convolution VS Correlation



$$F \circ I(x) = \sum_{i=-N}^{N} F(i) I(x+i)$$

[1] https://towardsdatascience.com/convolution-vs-correlation-af868b6b4fb5

Convolution VS Correlation

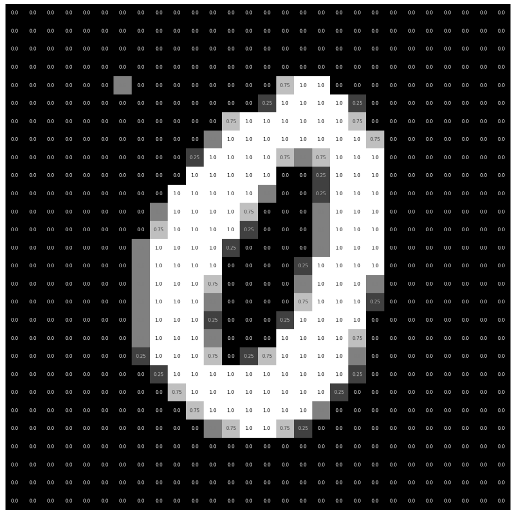


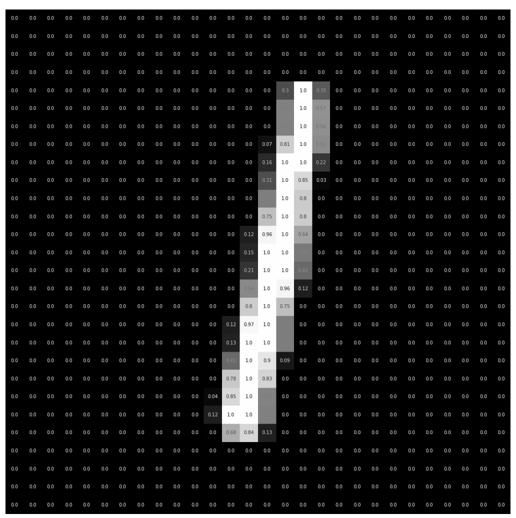
$$F \circ I(x,y) = \sum_{j=-N}^{N} \sum_{i=-N}^{N} F(i,j) I(x+i, y+j)$$

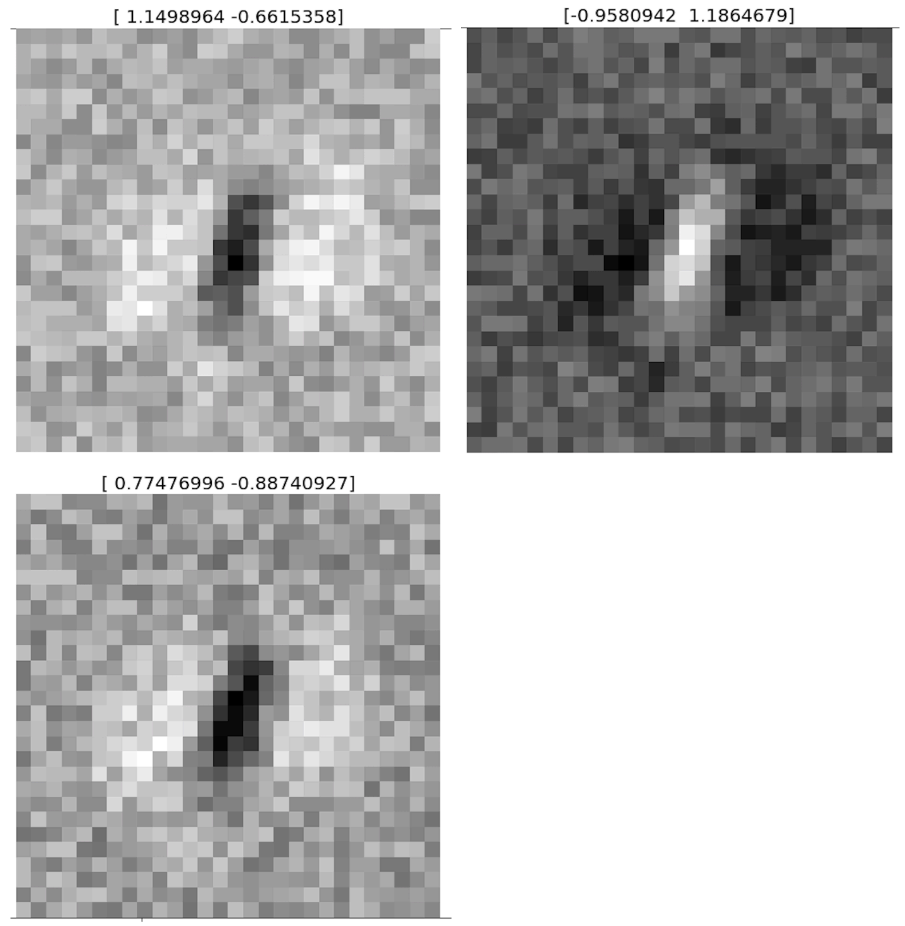
$$F * I(x,y) = \sum_{j=-N}^{N} \sum_{i=-N}^{N} F(i,j) I(x-i, y-j)$$

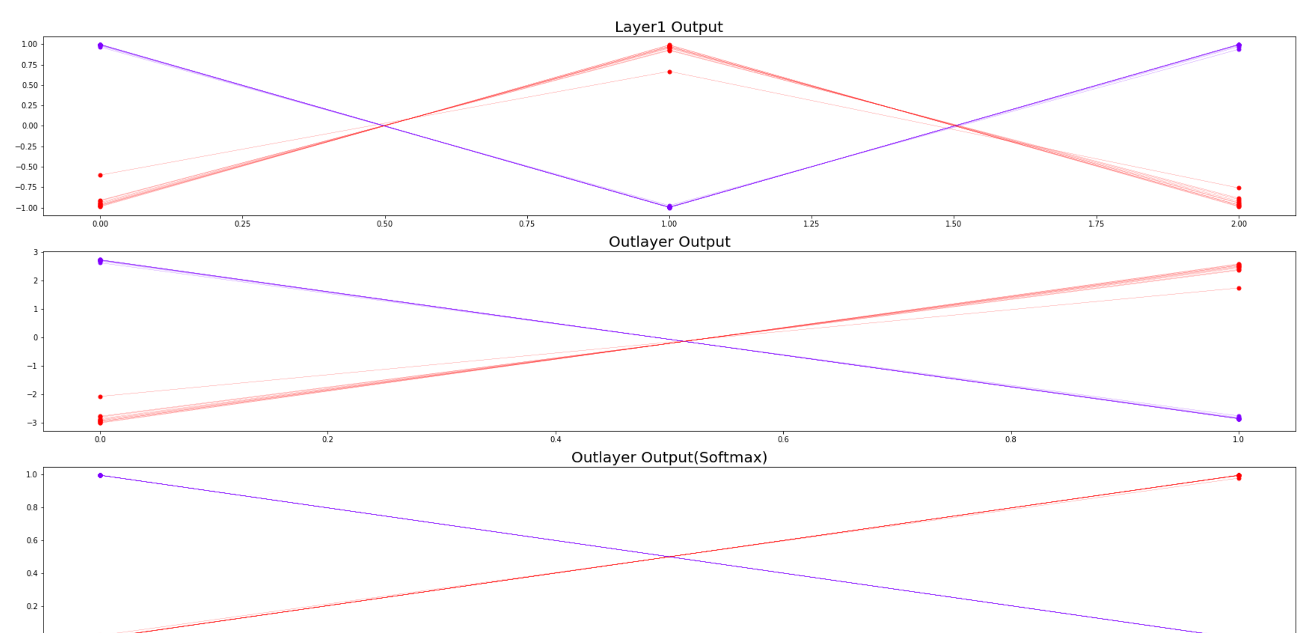
[1] https://towardsdatascience.com/convolution-vs-correlation-af868b6b4fb5

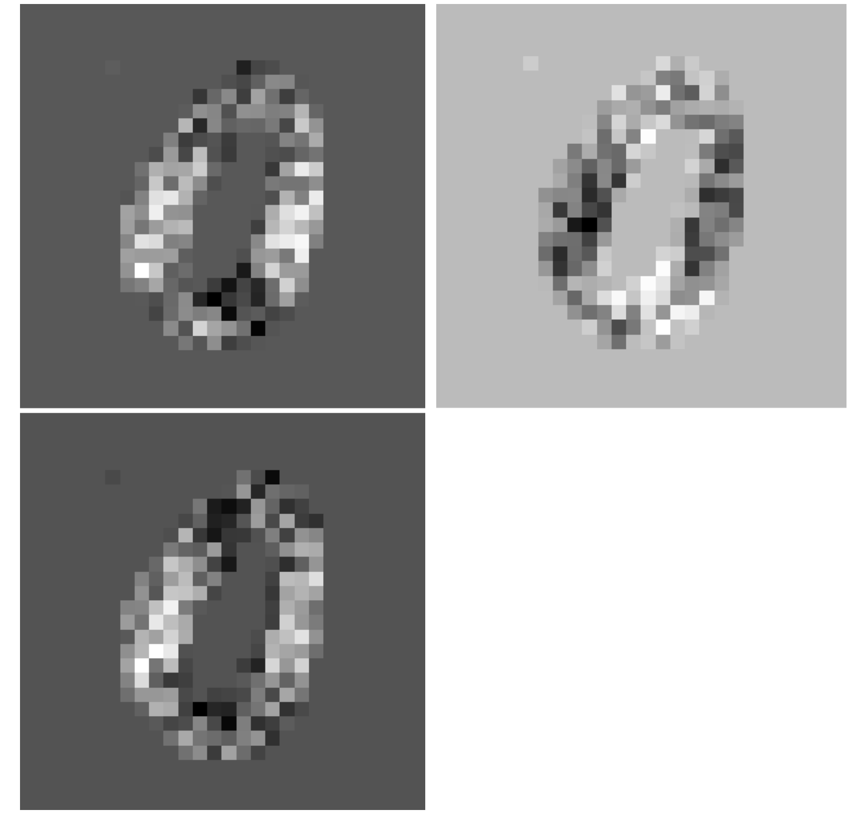
Convolution VS Correlation

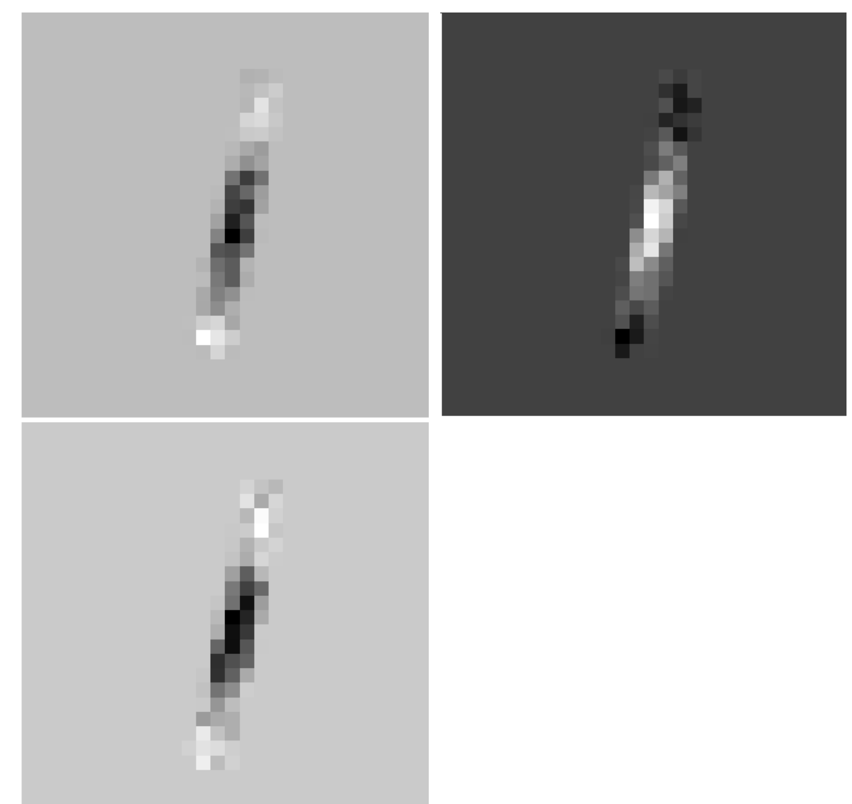




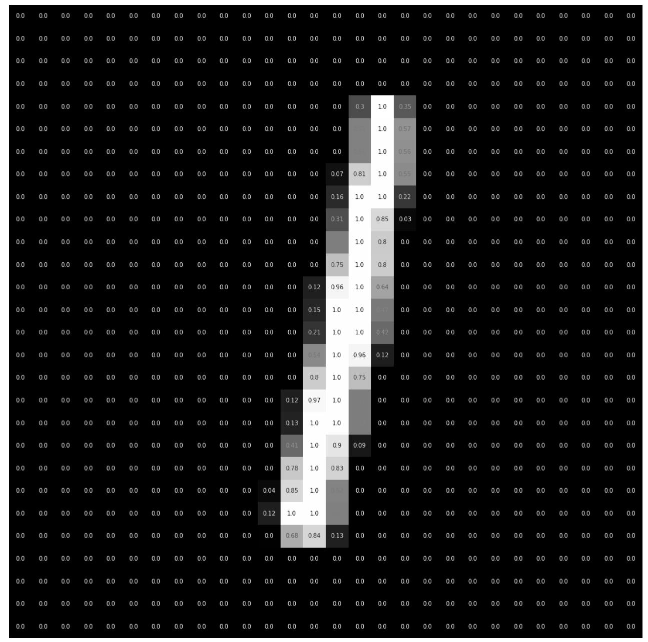


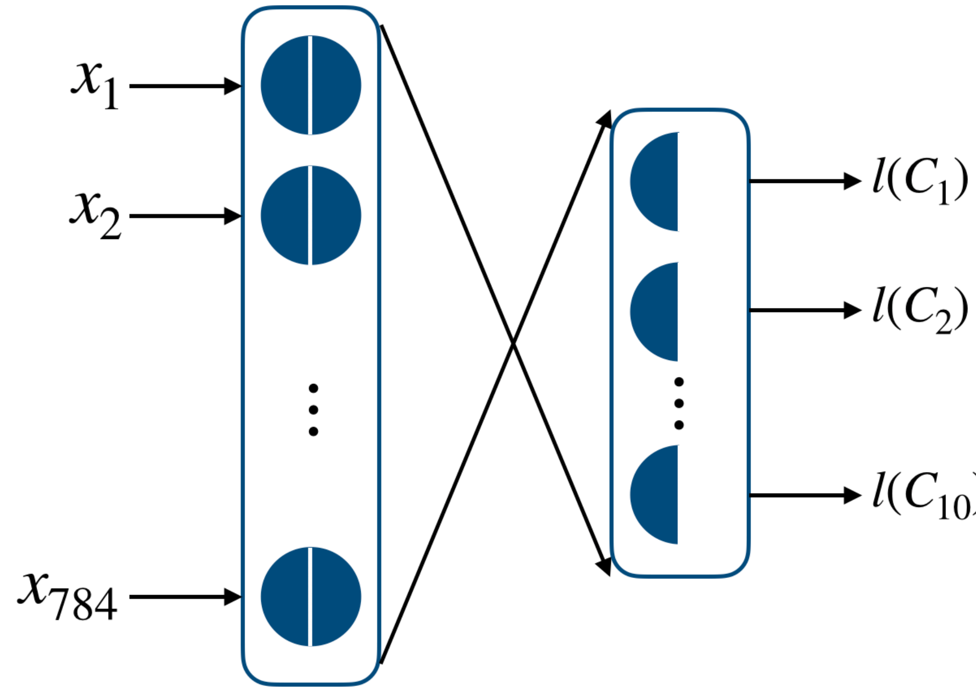






Forward/Backward Propagation

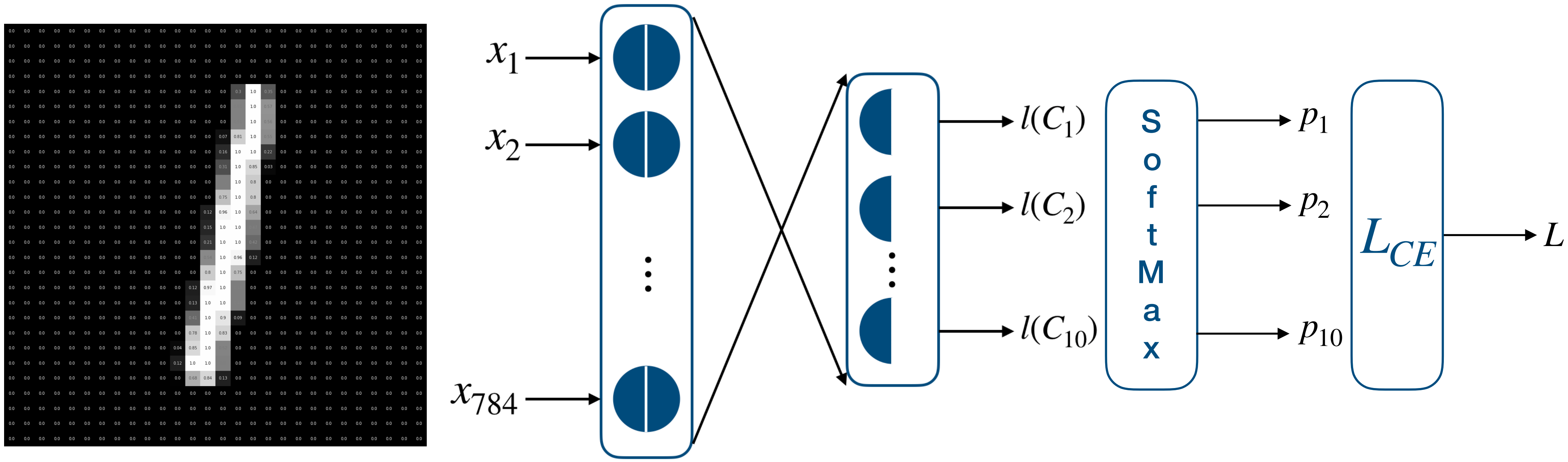




$$W\overrightarrow{x} + \overrightarrow{b} = \begin{pmatrix} \theta_{1,784} & \theta_{1,783} & \cdots & \theta_{1,1} \\ \theta_{2,784} & \theta_{2,783} & \cdots & \theta_{2,1} \\ \vdots & \vdots & \cdots & \vdots \\ \theta_{10,784} & \theta_{10,783} & \cdots & \theta_{10,1} \end{pmatrix} \begin{pmatrix} x_{784} \\ x_{783} \\ \vdots \\ x_1 \end{pmatrix} + \begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \\ \vdots \\ \theta_{10,0} \end{pmatrix}$$

$$= \begin{pmatrix} \theta_{1,784} & \theta_{1,783} & \cdots & \theta_{1,1} & \theta_{1,0} \\ \theta_{2,784} & \theta_{2,783} & \cdots & \theta_{2,1} & \theta_{2,0} \\ \vdots & \vdots & \cdots & \vdots & \\ \theta_{10,784} & \theta_{10,783} & \cdots & \theta_{10,1} & \theta_{10,0} \end{pmatrix} \begin{pmatrix} x_{784} \\ x_{783} \\ \vdots \\ x_1 \\ 1 \end{pmatrix}$$

Forward/Backward Propagation

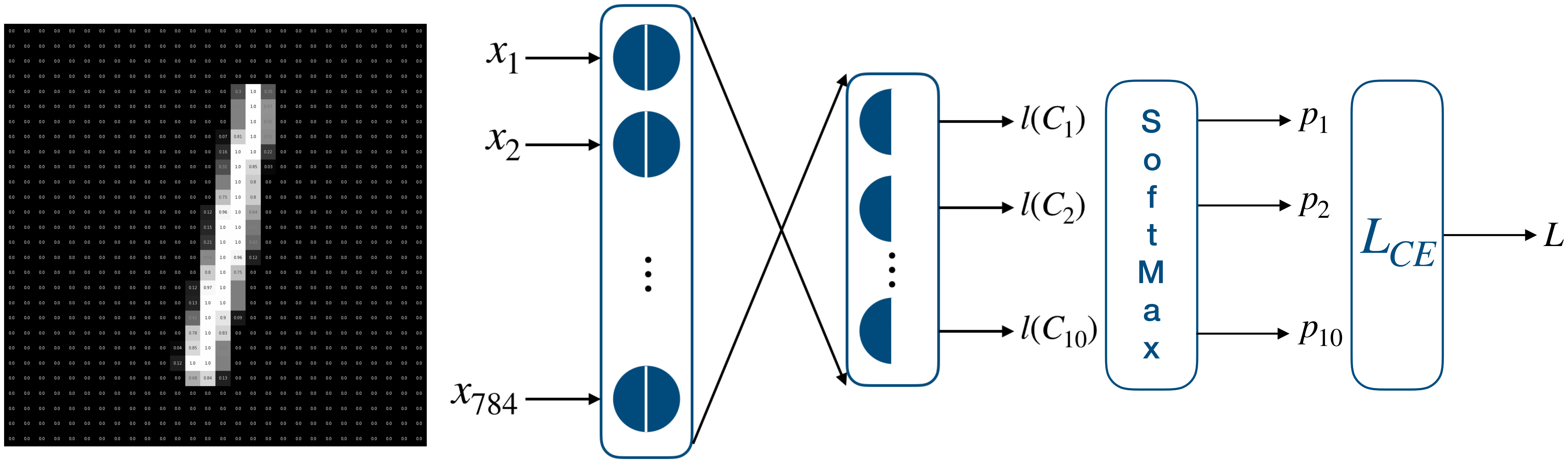


$$\begin{pmatrix} \theta_{1,784} & \theta_{1,783} & \cdots & \theta_{1,1} \\ \theta_{2,784} & \theta_{2,783} & \cdots & \theta_{2,1} \\ \vdots & \vdots & \cdots & \vdots \\ \theta_{10,784} & \theta_{10,783} & \cdots & \theta_{10,1} \end{pmatrix} \begin{pmatrix} x_{784} \\ x_{783} \\ \vdots \\ x_1 \end{pmatrix} + \begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \\ \vdots \\ \theta_{10,0} \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_{10} \end{pmatrix}$$

$$\begin{pmatrix} \frac{e^{(l_1)}}{\sum_k e^{(l_k)}} \\ \frac{e^{(l_2)}}{\sum_k e^{(l_k)}} \\ \vdots \\ \frac{e^{(l_{10})}}{\sum_k e^{(l_k)}} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{10} \end{pmatrix}$$

$$L_{CE} = -log(p_i)$$

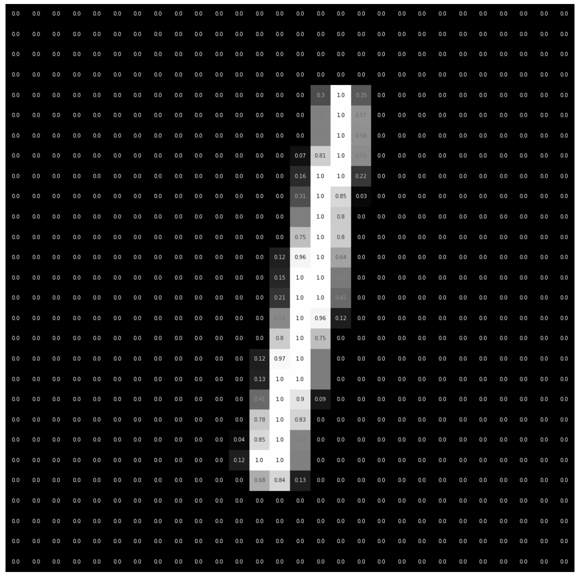
Forward/Backward Propagation

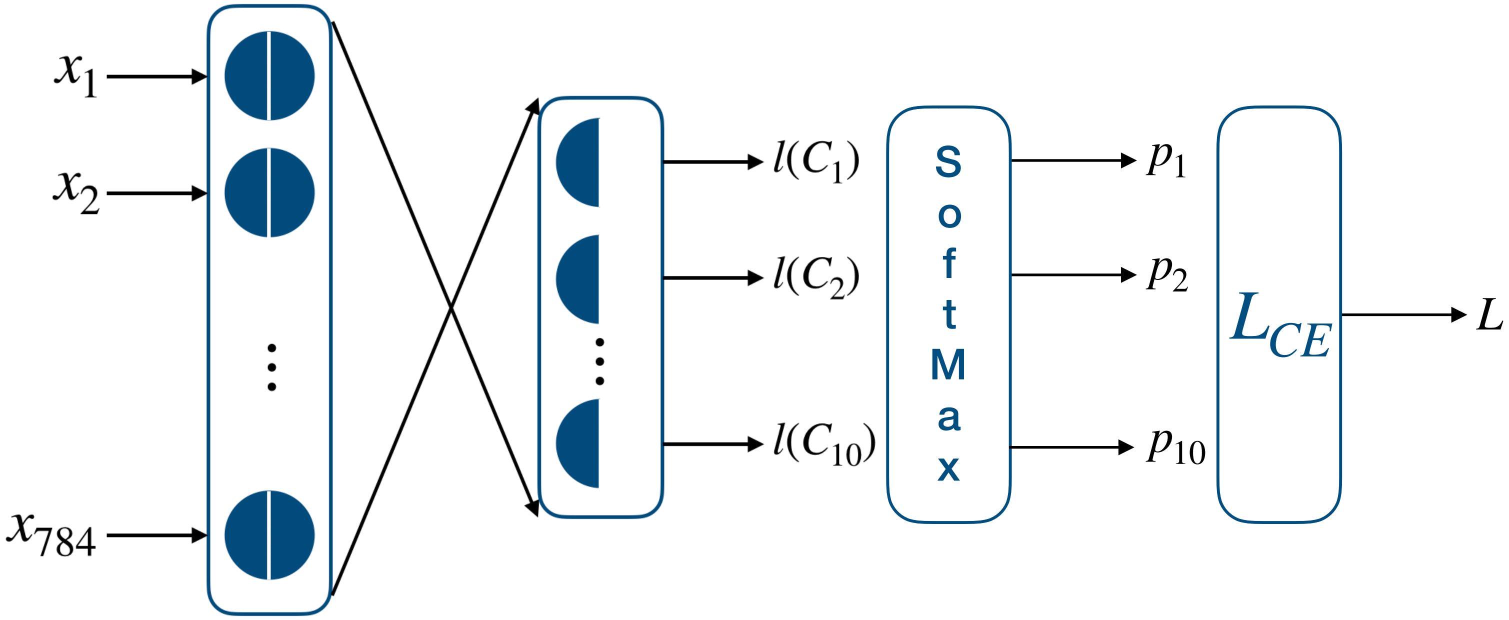


$$\begin{pmatrix} \theta_{1,784} & \theta_{1,783} & \cdots & \theta_{1,1} \\ \theta_{2,784} & \theta_{2,783} & \cdots & \theta_{2,1} \\ \vdots & \vdots & \cdots & \vdots \\ \theta_{10,784} & \theta_{10,783} & \cdots & \theta_{10,1} \end{pmatrix} \begin{pmatrix} x_{784} \\ x_{783} \\ \vdots \\ x_1 \end{pmatrix} + \begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \\ \vdots \\ \theta_{10,0} \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_{10} \end{pmatrix}$$

$$\frac{\partial l_\alpha}{\partial \overrightarrow{\theta_\alpha}} = \begin{pmatrix} \frac{\partial l_\alpha}{\partial \theta_{\alpha,784}} & \frac{\partial l_\alpha}{\partial \theta_{\alpha,783}} & \cdots & \frac{\partial l_\alpha}{\partial \theta_{\alpha,781}} \end{pmatrix} = (x_{784} \quad x_{783} \quad \cdots \quad x_1) \qquad \frac{\partial l_\alpha}{\partial \theta_{\alpha,0}} = 1$$

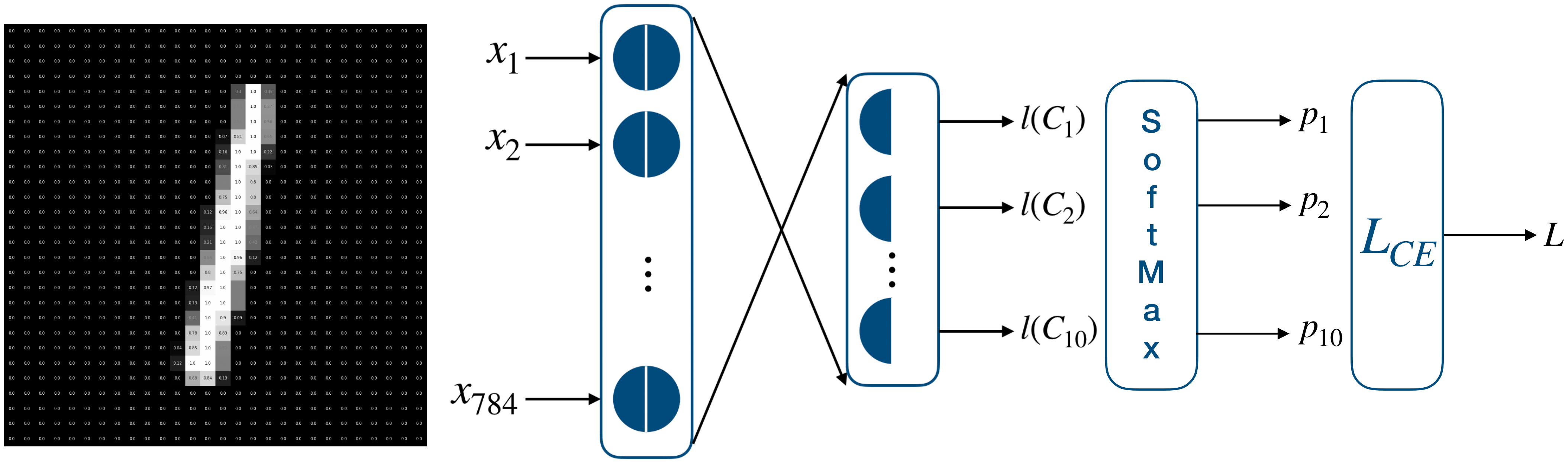
Forward/Backward Propagation





$$\begin{pmatrix} \frac{e^{(l_1)}}{\sum_k e^{(l_k)}} \\ \frac{e^{(l_2)}}{\sum_k e^{(l_k)}} \\ \vdots \\ \frac{e^{(l_{10})}}{\sum_k e^{(l_k)}} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{10} \end{pmatrix}$$

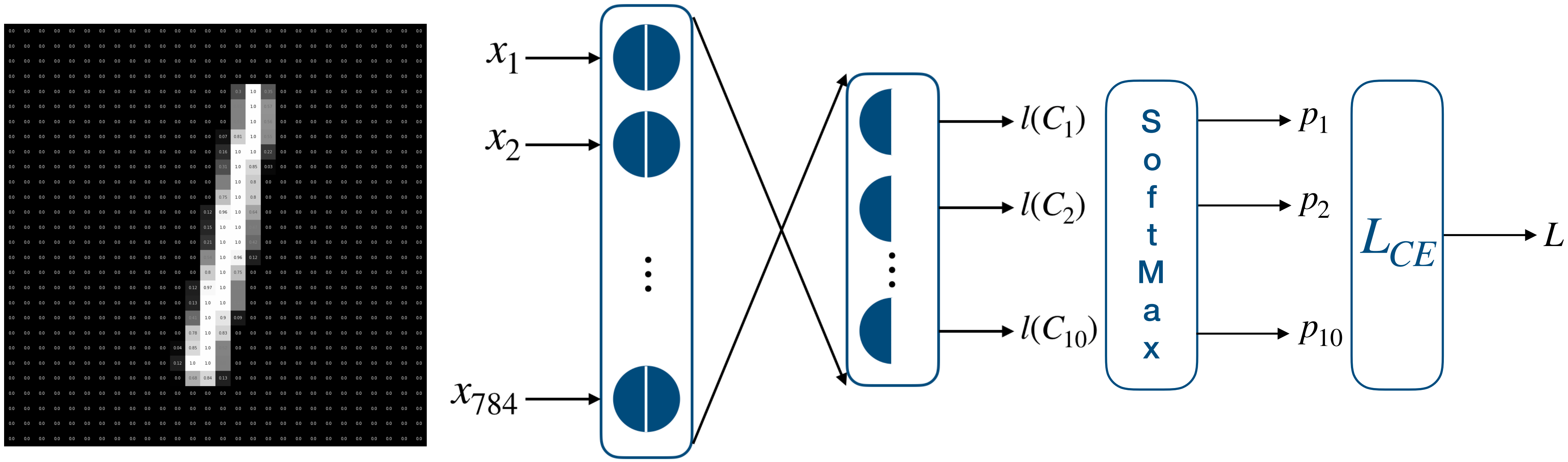
Forward/Backward Propagation



$$\begin{pmatrix} \frac{e^{(l_1)}}{\sum_k e^{(l_k)}} \\ \frac{e^{(l_2)}}{\sum_k e^{(l_k)}} \\ \vdots \\ \frac{e^{(l_{10})}}{\sum_k e^{(l_k)}} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{10} \end{pmatrix}$$

$$\frac{\partial \overrightarrow{p}}{\partial l_i} = \begin{pmatrix} \frac{\partial p_1}{\partial l_i} \\ \frac{\partial p_2}{\partial l_i} \\ \vdots \\ \frac{\partial p_i}{\partial l_i} \\ \vdots \\ \frac{\partial p_{10}}{\partial l_i} \end{pmatrix} = \begin{pmatrix} -p_i p_1 \\ -p_i p_2 \\ \vdots \\ p_i(1-p_i) \\ \vdots \\ -p_i p_{10} \end{pmatrix}$$

Forward/Backward Propagation



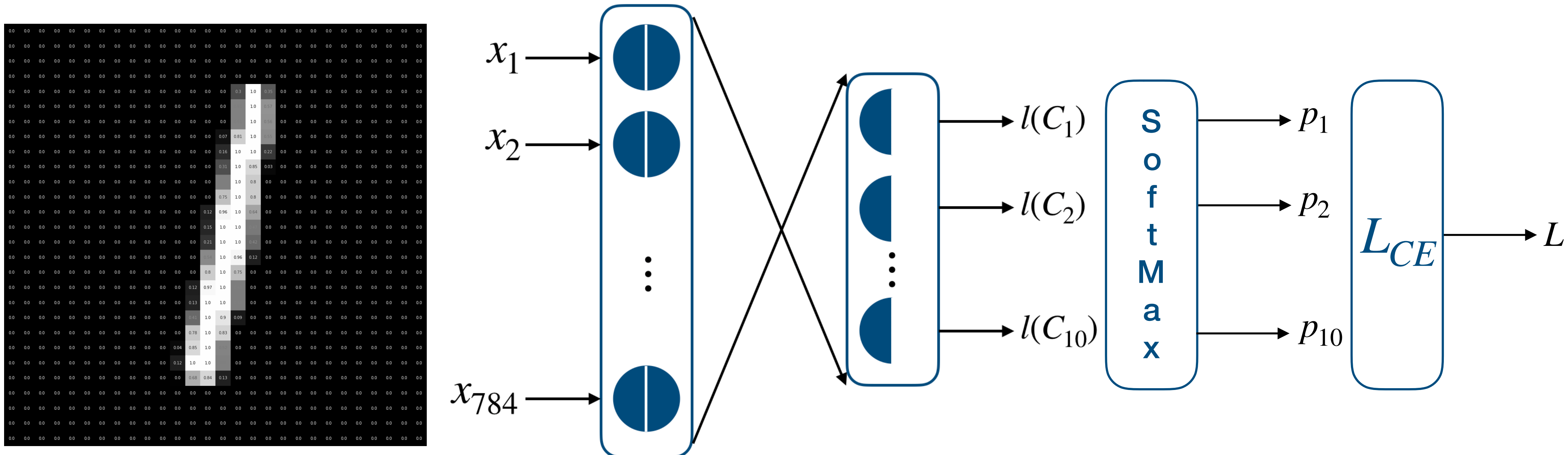
$$L_{CE} = -\,log(p_i)$$

$$\frac{\partial L_{CE}}{\partial p_\alpha} = \begin{cases} -\frac{1}{p_i}, & where\ \alpha = i \\ 0, & otherwise \end{cases}$$

$$\frac{\partial L_{CE}}{\partial \overrightarrow{p}} = \begin{pmatrix} \frac{\partial L_{CE}}{\partial p_1} & \frac{\partial L_{CE}}{\partial p_2} & \dots & \frac{\partial L_{CE}}{\partial p_i} & \dots & \frac{\partial L_{CE}}{\partial p_{10}} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & \dots & -\frac{1}{p_i} & \dots & 0 \end{pmatrix}$$

Forward/Backward Propagation



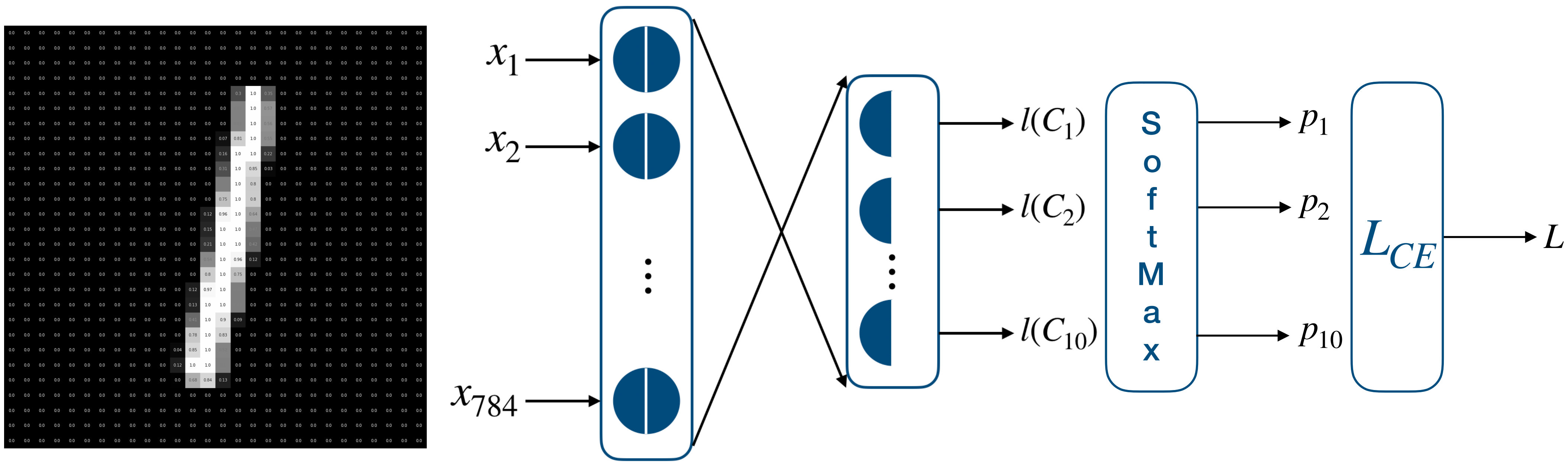
$$\frac{\partial L_{CE}}{\partial \vec{p}} = \begin{pmatrix} 0 & 0 & \dots & -\frac{1}{p_i} & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{p}}{\partial l_i} = \begin{pmatrix} \frac{\partial p_1}{\partial l_i} \\ \frac{\partial p_2}{\partial l_i} \\ \vdots \\ \frac{\partial p_i}{\partial l_i} \\ \vdots \\ \frac{\partial p_{10}}{\partial l_i} \end{pmatrix} = \begin{pmatrix} -p_i p_1 \\ -p_i p_2 \\ \vdots \\ p_i(1-p_i) \\ \vdots \\ -p_i p_{10} \end{pmatrix}$$

$$\begin{aligned} \frac{\partial L_{CE}}{\partial l_1} &= \frac{\partial L_{CE}}{\partial \vec{p}} \frac{\partial \vec{p}}{\partial l_1} \\ &= \begin{pmatrix} 0 & 0 & \dots & -\frac{1}{p_i} & \dots & 0 \end{pmatrix} \begin{pmatrix} p_1(1-p_1) \\ -p_1 p_2 \\ \vdots \\ -p_1 p_i \\ \vdots \\ -p_1 p_{10} \end{pmatrix} \\ &= p_1 \end{aligned}$$

$$\begin{aligned} \frac{\partial L_{CE}}{\partial l_i} &= \frac{\partial L_{CE}}{\partial \vec{p}} \frac{\partial \vec{p}}{\partial l_i} \\ &= \begin{pmatrix} 0 & 0 & \dots & -\frac{1}{p_i} & \dots & 0 \end{pmatrix} \begin{pmatrix} -p_i p_1 \\ -p_i p_2 \\ \vdots \\ p_i(1-p_i) \\ \vdots \\ -p_1 p_{10} \end{pmatrix} \\ &= p_i - 1 \end{aligned}$$

Forward/Backward Propagation



$$\frac{\partial l_\alpha}{\partial \overrightarrow{\theta_\alpha}} = \begin{pmatrix} \frac{\partial l_\alpha}{\partial \theta_{\alpha,784}} & \frac{\partial l_\alpha}{\partial \theta_{\alpha,783}} & \cdots & \frac{\partial l_\alpha}{\partial \theta_{\alpha,781}} \end{pmatrix} = (x_{784} \quad x_{783} \quad \cdots \quad x_1) \qquad \frac{\partial l_1}{\partial \theta_{1,0}} = 1$$

$$\frac{\partial L_{CE}}{\partial \overrightarrow{\theta_1}} = \frac{\partial L_{CE}}{\partial \overrightarrow{p}} \frac{\partial \overrightarrow{p}}{\partial l_1} \frac{\partial l_1}{\partial \overrightarrow{\theta_1}}$$

$$= p_1 \, (x_{784} \quad x_{783} \quad \cdots \quad x_1)$$

$$\frac{\partial L_{CE}}{\partial \overrightarrow{\theta_i}} = \frac{\partial L_{CE}}{\partial \overrightarrow{p}} \frac{\partial \overrightarrow{p}}{\partial l_i} \frac{\partial l_i}{\partial \overrightarrow{\theta_i}}$$

$$= (p_i - 1)(x_{784} \quad x_{783} \quad \cdots \quad x_1)$$

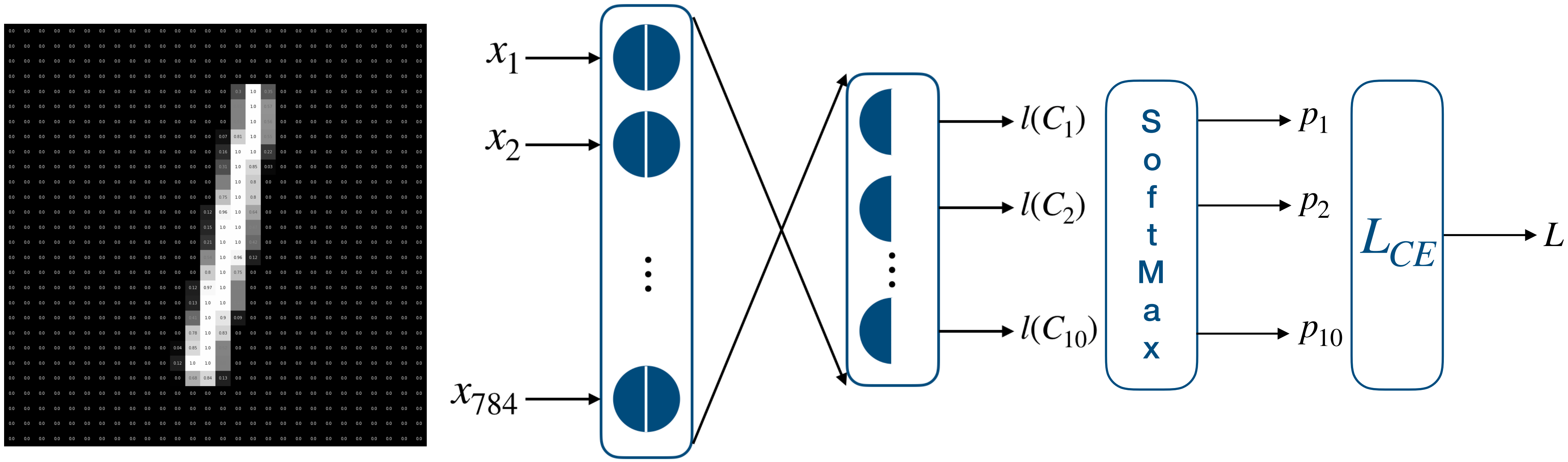
$$\frac{\partial L_{CE}}{\partial \theta_{1,0}} = \frac{\partial L_{CE}}{\partial \overrightarrow{p}} \frac{\partial \overrightarrow{p}}{\partial l_1} \frac{\partial l_1}{\partial \theta_{1,0}}$$

$$= p_1$$

$$\frac{\partial L_{CE}}{\partial \theta_{i,0}} = \frac{\partial L_{CE}}{\partial \overrightarrow{p}} \frac{\partial \overrightarrow{p}}{\partial l_i} \frac{\partial l_i}{\partial \theta_{i,0}}$$

$$= p_i - 1$$

Forward/Backward Propagation



$$\overrightarrow{\theta_1} = \overrightarrow{\theta_1} - \alpha * \frac{\partial L_{CE}}{\partial \overrightarrow{\theta_1}}$$

$$= \overrightarrow{\theta_1} - \alpha * p_1 \begin{pmatrix} x_{784} & x_{783} & \cdots & x_1 \end{pmatrix}$$

$$\overrightarrow{\theta_i} = \overrightarrow{\theta_i} - \alpha * \frac{\partial L_{CE}}{\partial \overrightarrow{\theta_i}}$$

$$= \overrightarrow{\theta_i} - \alpha * (p_i - 1) \begin{pmatrix} x_{784} & x_{783} & \cdots & x_1 \end{pmatrix}$$

$$\theta_{1,0} = \theta_{1,0} - \alpha * \frac{\partial L_{CE}}{\partial \theta_{1,0}}$$

$$= \theta_{1,0} - \alpha * p_1$$

$$\theta_{i,0} = \theta_{i,0} - \alpha * \frac{\partial L_{CE}}{\partial \theta_{i,0}}$$

$$= \theta_{1,0} - \alpha * (p_i - 1)$$

MNIST Filter