Lab1 – Understand NN and Training Process

Advisor：Tsai, Chia-Chi

TA：賴姿伶

Outline

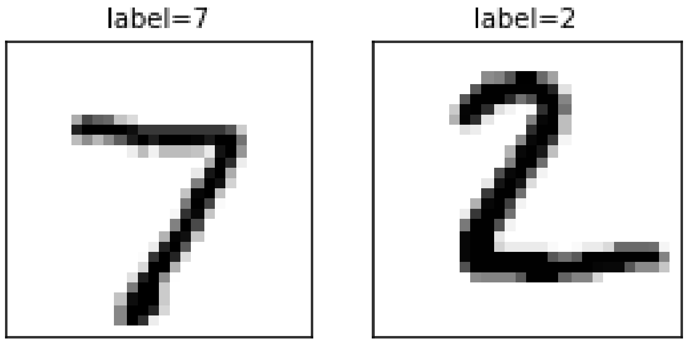
1. Lab1 tasks
2. Google Colab
3. How NN works
4. MNIST dataset

Outline

1. Lab1 tasks
2. Google Colab
3. How NN works
4. MNIST dataset

1. Lab1 tasks

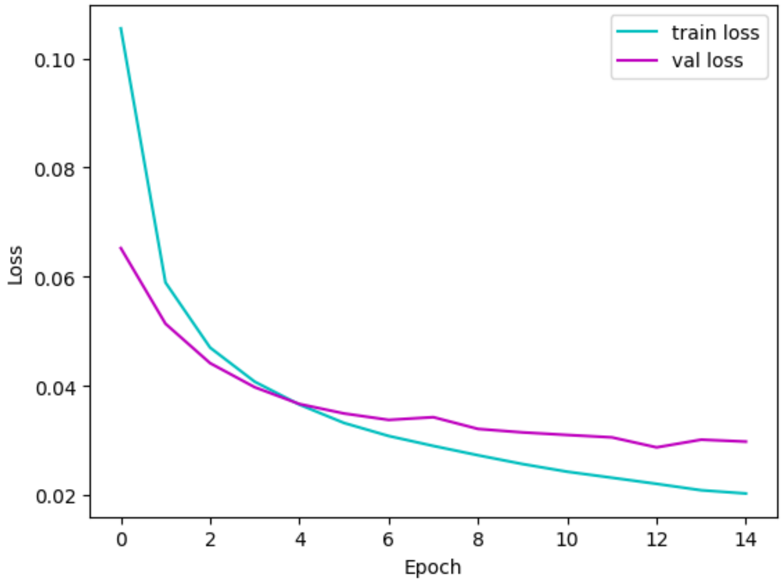
- Implement the following layers as a python function (both forward and backward propagation)
  - Inner-product layer (10%)
  - Activation layer (Sigmoid or Rectified) (10%)
  - Softmax layer (10%)
- Implement training and testing process
  - Included cross-validation (5%)
  - Use cross-entropy as loss function (5%)
- Build neural network to solve the MNIST classification problem
  - At least one hidden layer neural network (cannot use convolutional layer) (10%)
  - Accuracy must > 90%
- Print test and val accuracy. Plot epoch-train accuracy, epoch-val accuracy, epoch-train loss, epoch-val loss (10%)
- Report (40%)

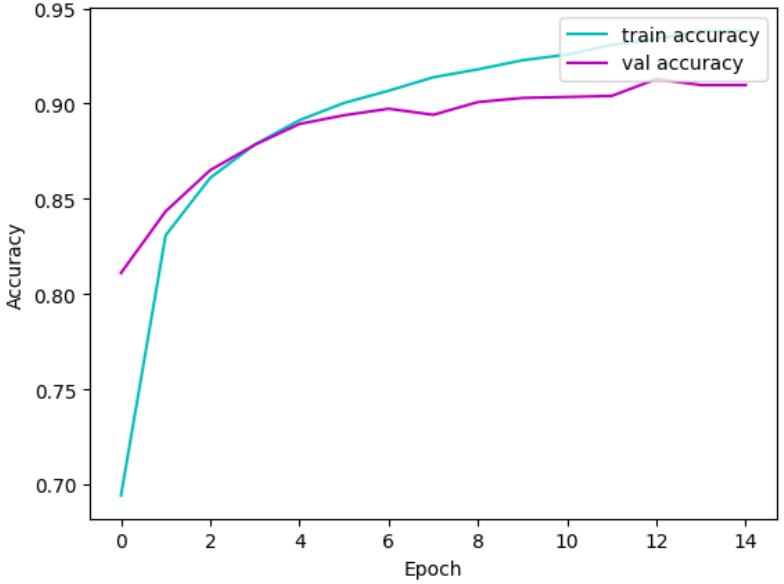


Display your result

- Print val accuracy of each epoch.
- Plot epoch-accuracy and epoch-loss.
- Print test accuracy.

```
[ Train | 005/015 ] loss = 0.03656, acc = 0.89131
[ Validation | 005/015 ] loss = 0.03667, acc = 0.88933
[ Train | 006/015 ] loss = 0.03320, acc = 0.90038
[ Validation | 006/015 ] loss = 0.03494, acc = 0.89392
[ Train | 007/015 ] loss = 0.03081, acc = 0.90681
[ Validation | 007/015 ] loss = 0.03376, acc = 0.89733
[ Train | 008/015 ] loss = 0.02899, acc = 0.91385
[ Validation | 008/015 ] loss = 0.03425, acc = 0.89417
[ Train | 009/015 ] loss = 0.02728, acc = 0.91802
[ Validation | 009/015 ] loss = 0.03211, acc = 0.90083
```





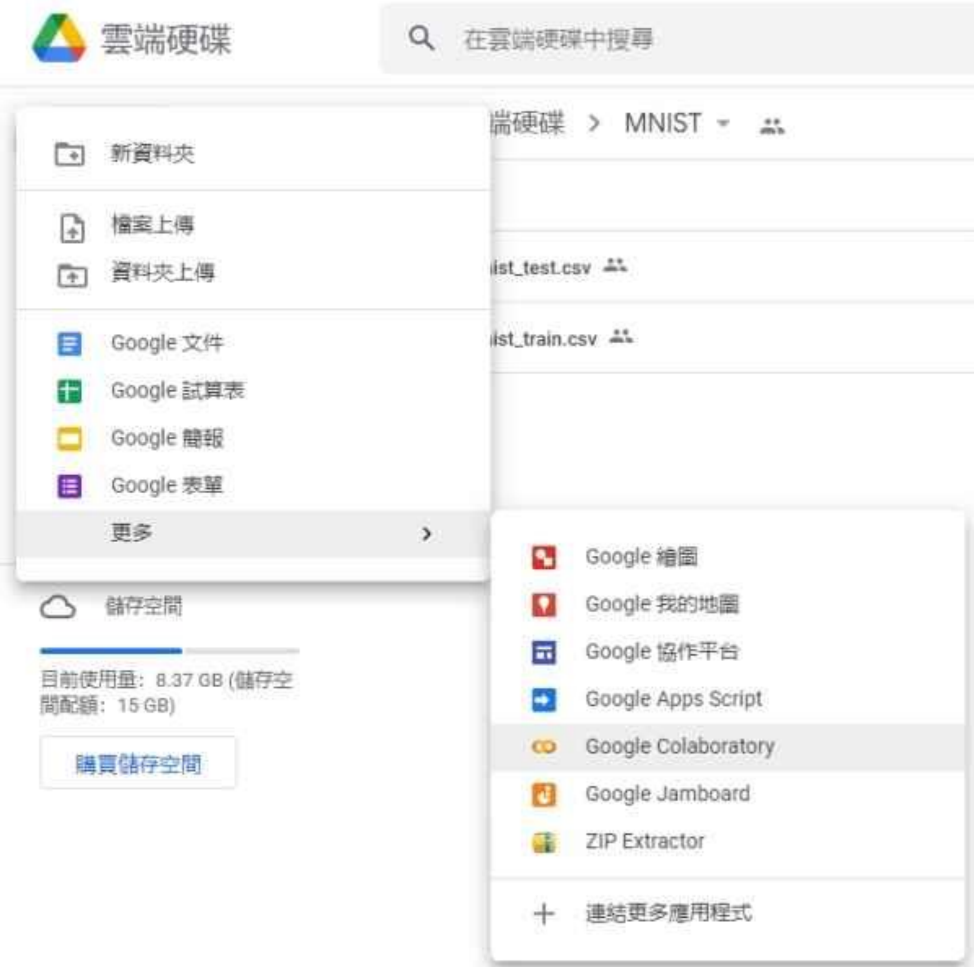

```
[ Test ] loss = 0.03162, acc = 0.90630
```

Outline

1. Lab1 tasks
2. Google Colab
3. How NN works
4. MNIST dataset

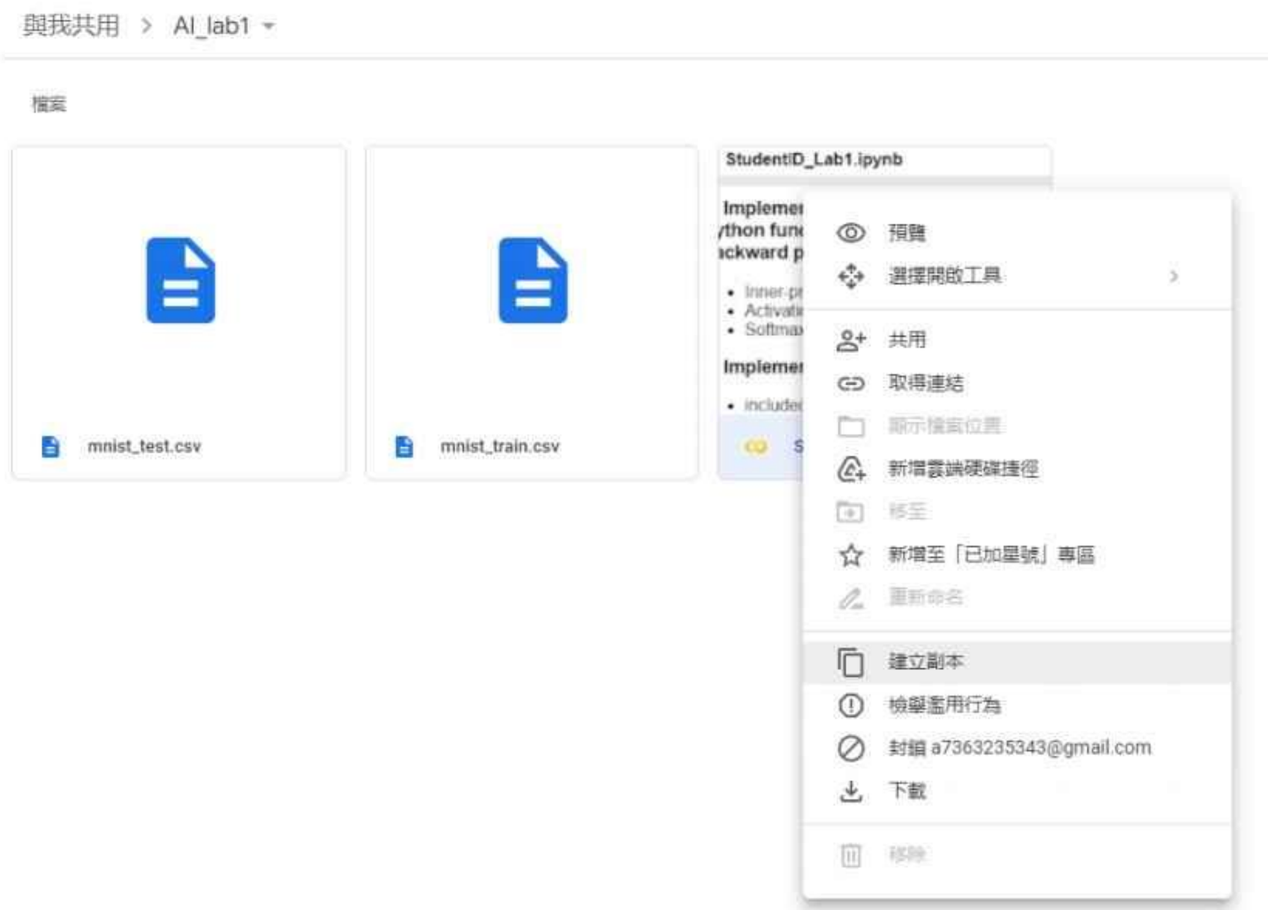
2. Google Colab

- Go to Google Drive and Create a Google Colaboratory file



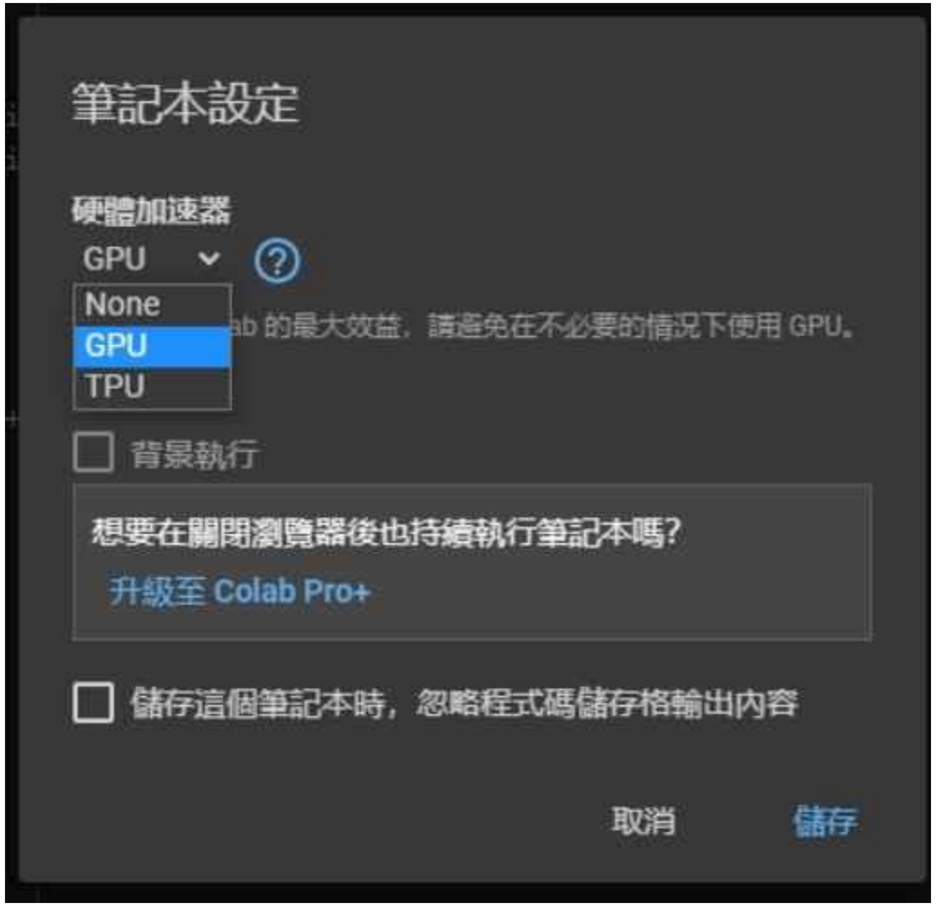
2. Google Colab

- 於sample code點選”在雲端硬碟中儲存副本”



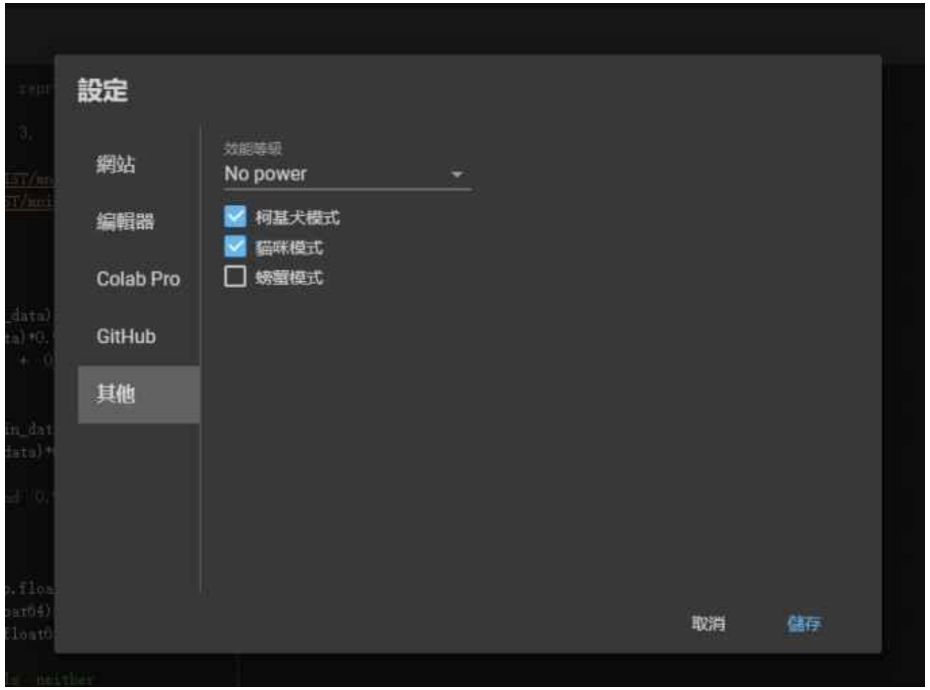
2. Google Colab

- Choose GPU as runtime (Default : CPU)



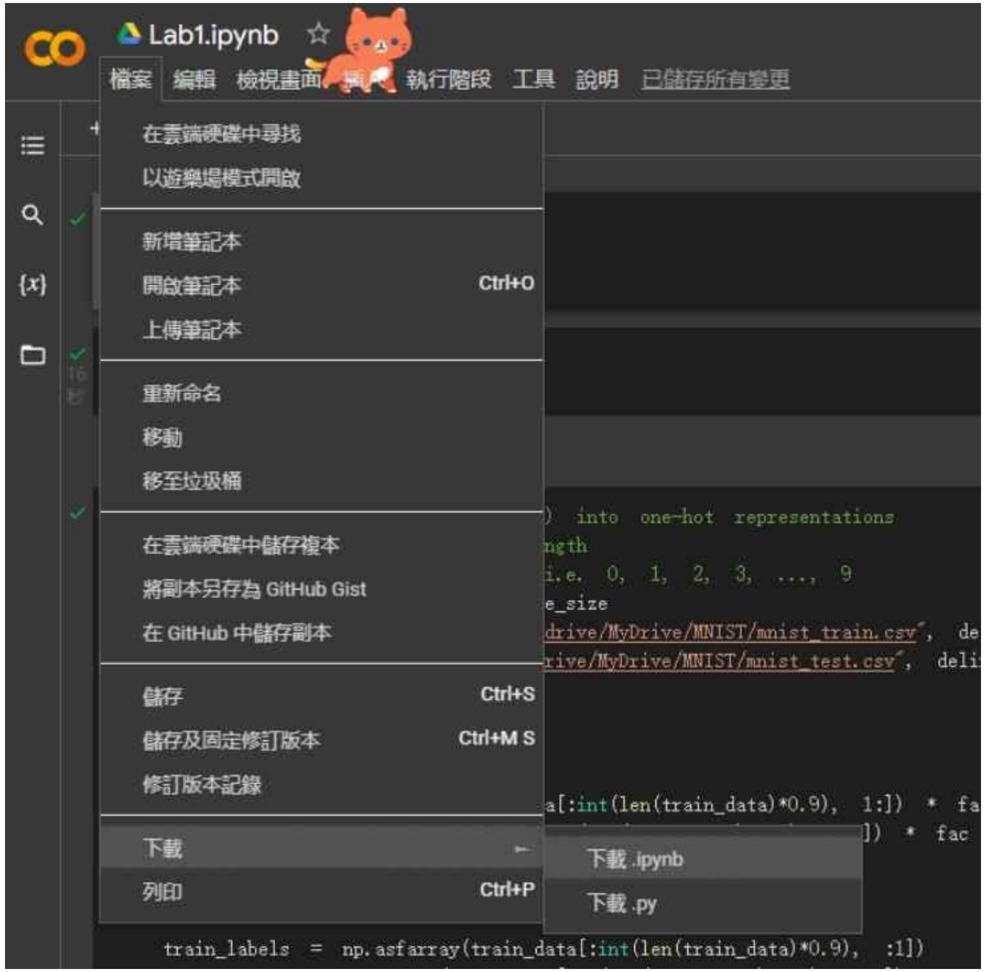
2. Google Colab

- Choose a pet



2. Google Colab

- Save as ipynb



2. Google Colab

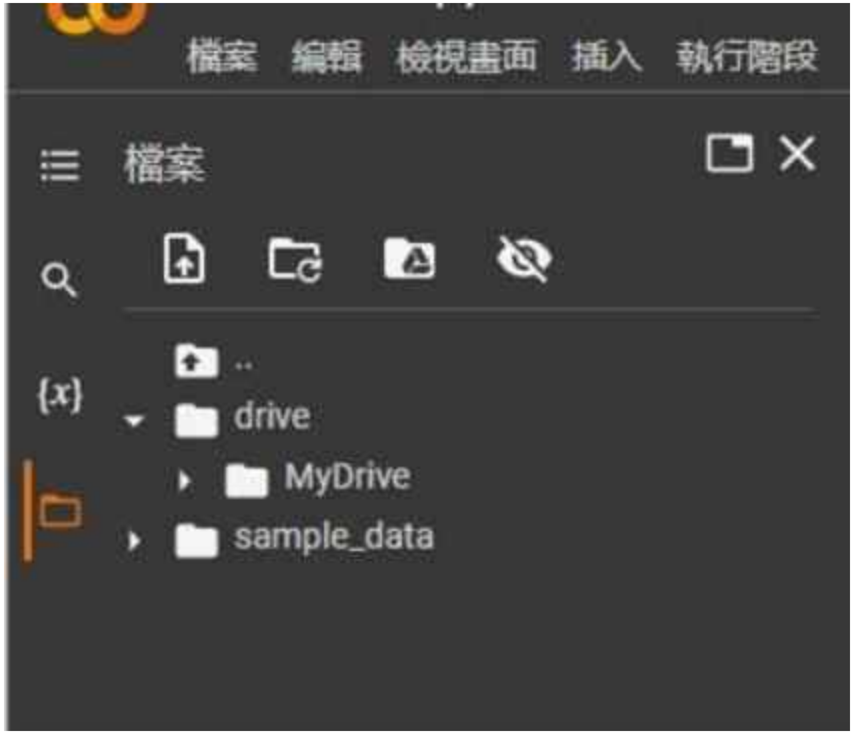
- 連結到雲端資料夾並設定當前路徑

```
[ ]  from google.colab import drive
     drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
[ ]  %cd /content/drive/MyDrive/lab1
```

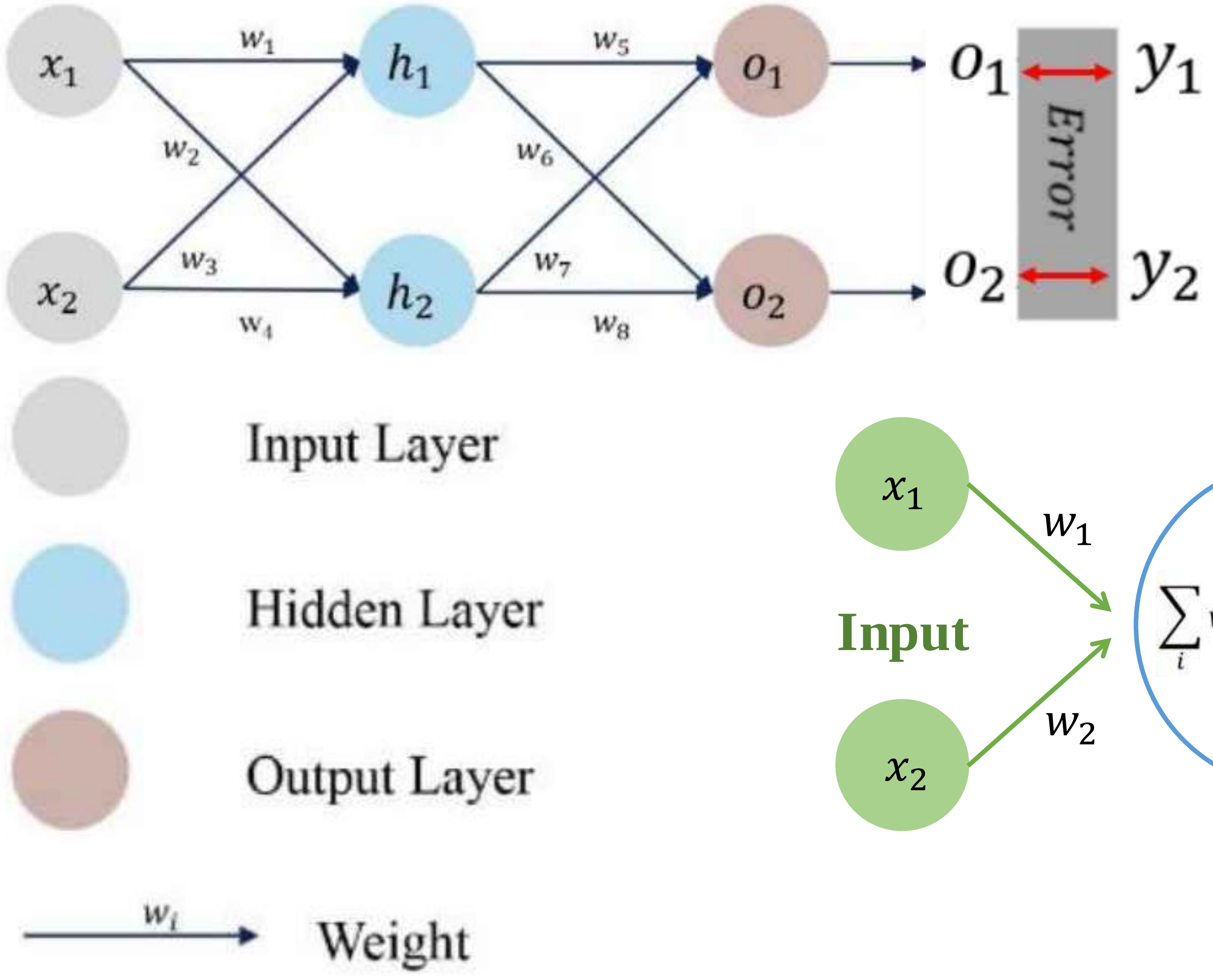
```
/content/drive/MyDrive/lab1
```



Outline

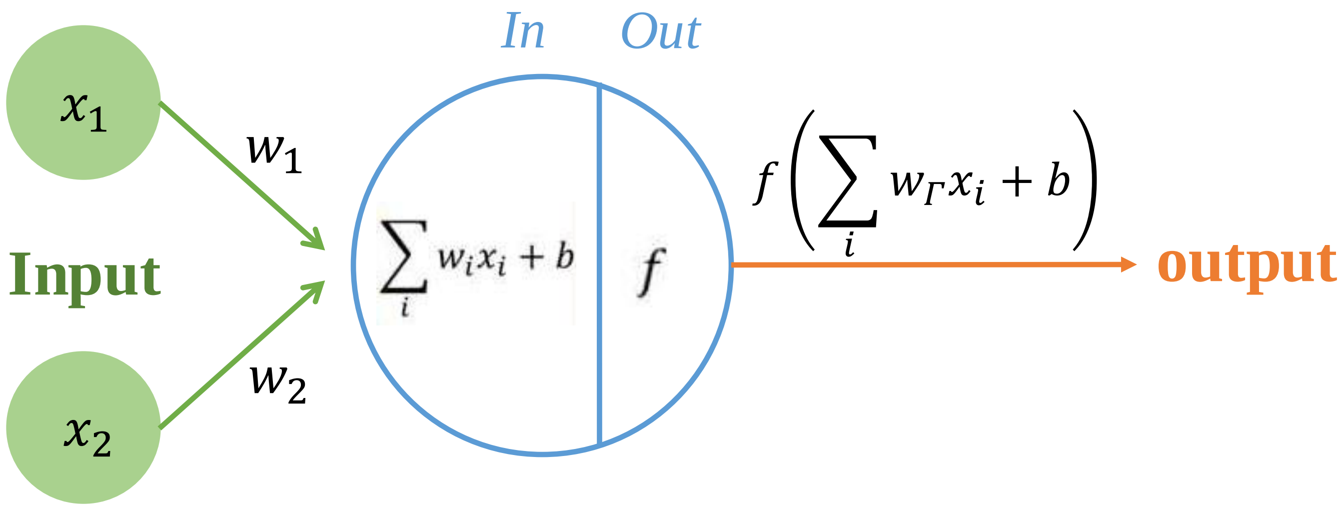
1. Lab1 tasks
2. Google Colab
3. How NN works
4. MNIST dataset

NN training



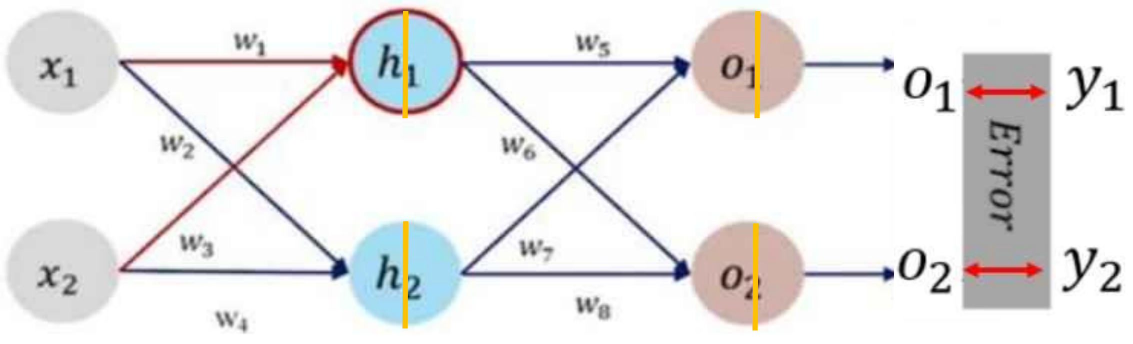
$$y = \sum_{i=1}^{N} w_i x_i + b$$

f *is activation function*
$f(x) = Sigmoid(x)$



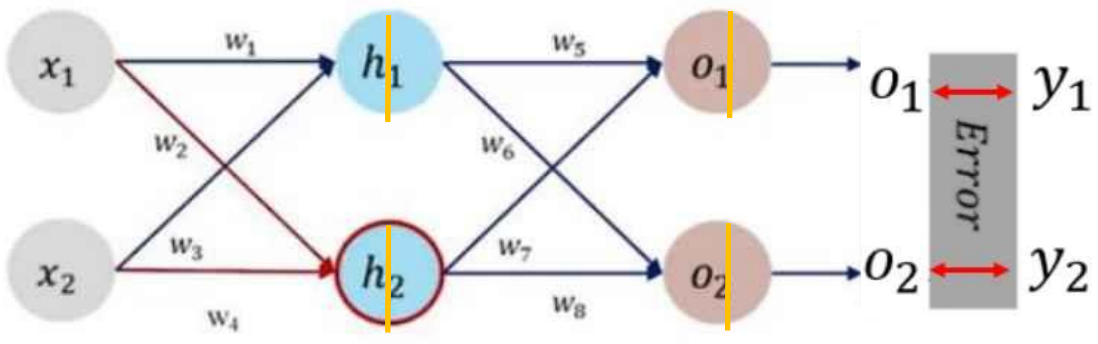
‼ Lab 1 use Softmax as the output layer

NN training – forward propagation



$$In_{h_1} = w_1 * x_1 + w_3 * x_2$$

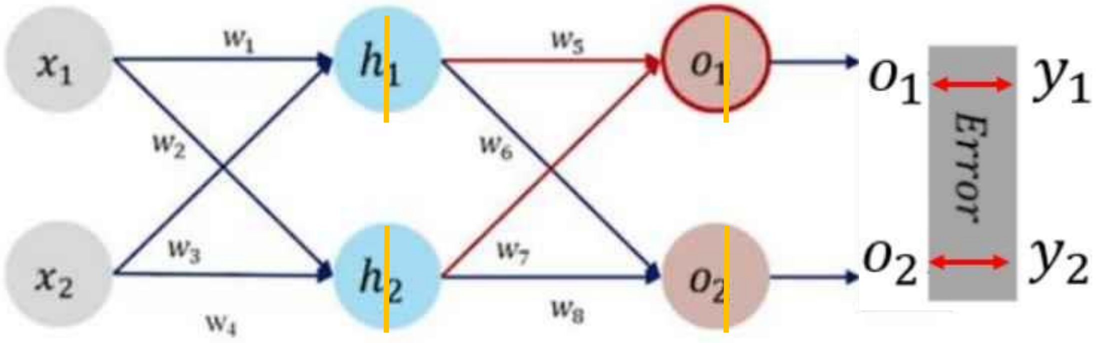
$$h_1 = Out_{h_1} = Sigmoid(In_{h_1})$$



$$In_{h_2} = w_2 * x_1 + w_4 * x_2$$

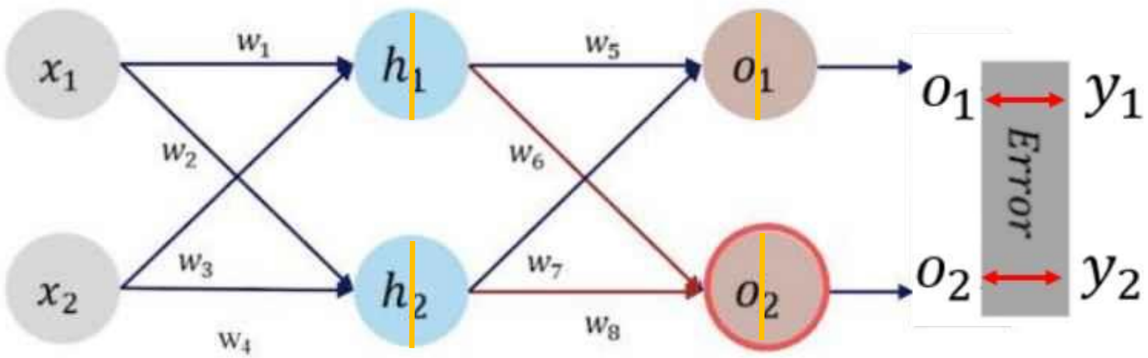
$$h_2 = Out_{h_2} = Sigmoid(In_{h_2})$$

NN training – forward propagation



$$In_{o_1} = w_5 * h_1 + w_7 * h_2$$

$$o_1 = Out_{o_1} = Sigmoid(In_{o_1})$$



$$In_{o_2} = w_6 * h_1 + w_8 * h_2$$

$$o_2 = Out_{o_2} = Sigmoid(In_{o_2})$$

MSE : $Error = \frac{1}{2} \sum_{i=1}^{2} (o_i - y_i)^2$

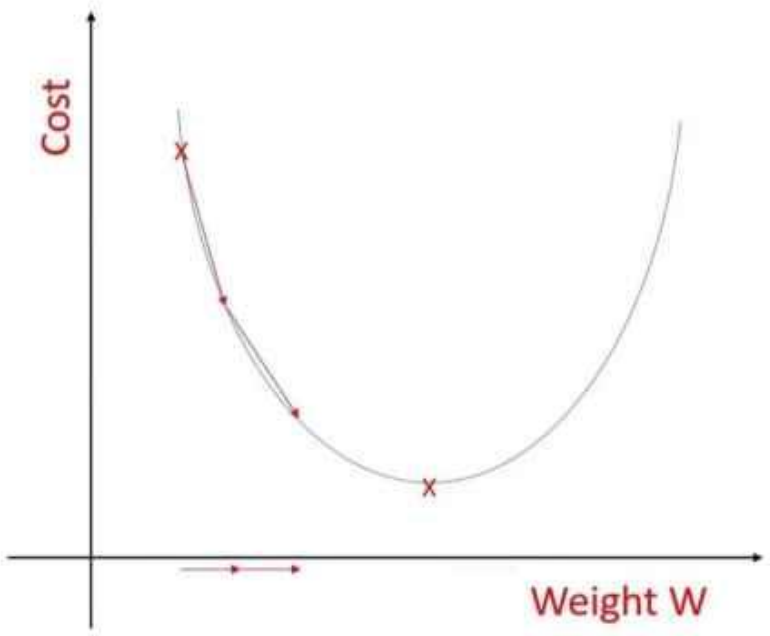
Gradient Descent

- Calculate the gradient of the loss function with respect to the weights, and adjust the weights along the gradient direction to reduce the loss.
- In the neural networks, there are only the inner-product layers have weights to be optimized.
- Using gradient descent to optimize parameters in inner-product layers.

$$\boldsymbol{W}^{new} = \boldsymbol{W}^{old} - \eta \nabla_{\boldsymbol{W}} E$$

$$\boldsymbol{b}^{new} = \boldsymbol{b}^{old} - \eta \nabla_{\boldsymbol{b}} E$$

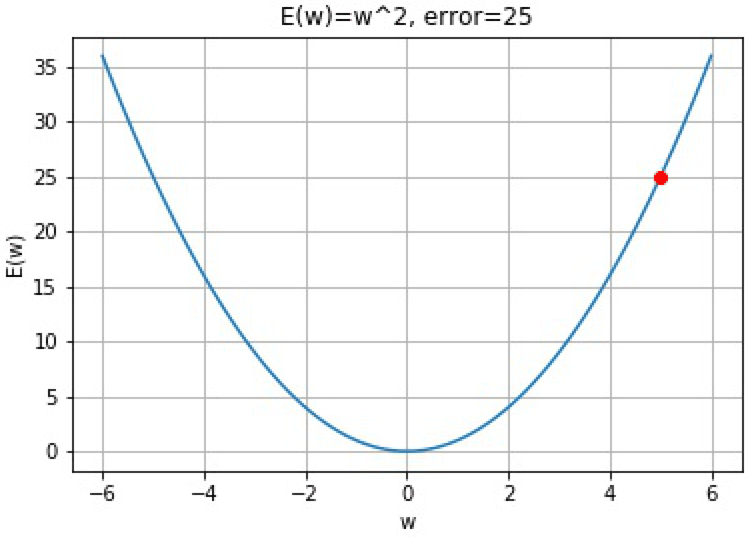
η is learning rate.



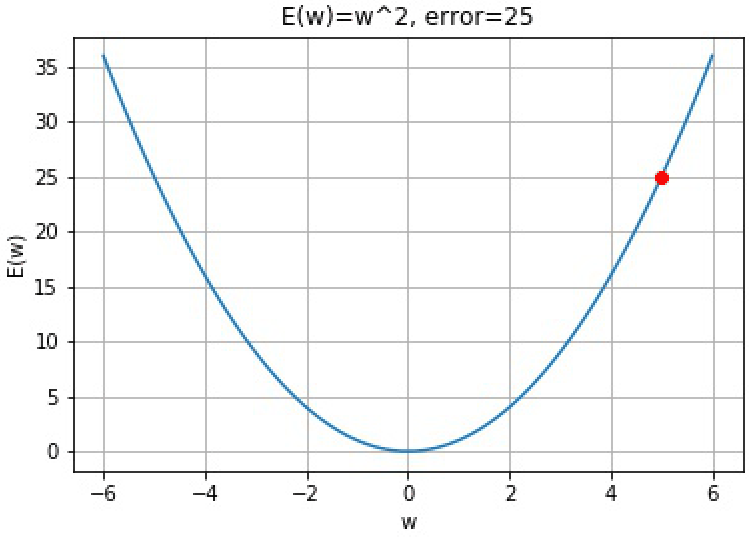
Gradient Descent

- Learning rate determines the step size.
- If the learning rate is too large, there will often be instability and easily to miss the best result.
- If the learning rate is too small, more iterations are needed to reach the minimum value.

Learning rate = 0.9



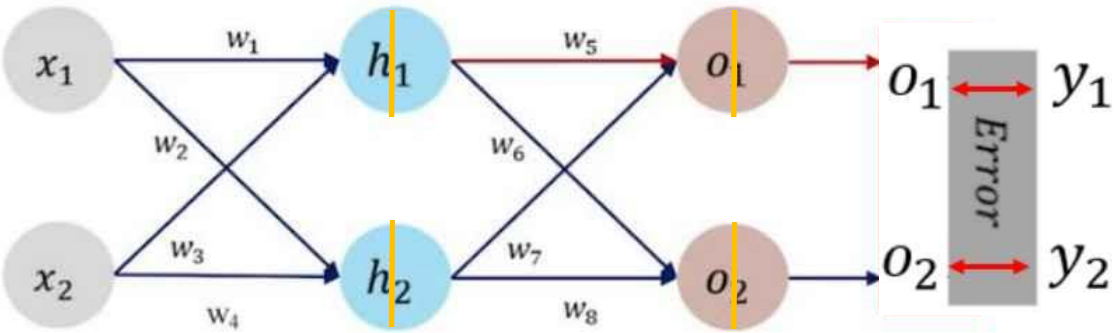
Learning rate = 0.1



NN training – backward propagation

- 計算w_5到w_8的梯度(誤差對每個權重的變化)

(以w_5 為例)



$$\delta_5 = \frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial o_1} * \frac{\partial o_1}{\partial In_{o_1}} * \frac{\partial In_{o_1}}{\partial w_5}$$

$$\frac{\partial Error}{\partial o_1} = o_1 - y_1 \quad \Leftarrow \quad Error = \frac{1}{2}\sum_{i=1}^{2}(o_i - y_i)^2$$

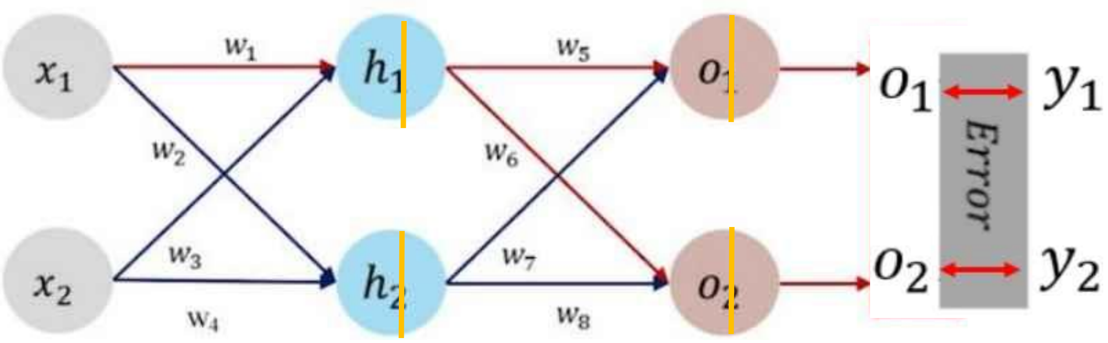
$$\frac{\partial o_1}{\partial In_{o_1}} = o_1 * (1 - o_1) \quad \Leftarrow \quad o_1 = Out_{o_1} = Sigmoid(In_{o_1})$$

$$\frac{\partial In_{o_1}}{\partial w_5} = h_1 \quad \Leftarrow \quad In_{o_1} = w_5 * h_1 + w_7 * h_2$$

NN training – backward propagation

- 計算w_1到w_4的梯度(誤差對每個權重的變化)

(以w_1 為例)



$$\delta_1 = \frac{\partial Error}{\partial w_1} = \frac{\partial Error}{\partial o_1} \cdot \frac{\partial o_1}{\partial w_1} + \frac{\partial Error}{\partial o_2} \cdot \frac{\partial o_2}{\partial w_1}$$

$$= \frac{\partial Error}{\partial o_1} \cdot \frac{\partial o_1}{\partial In_{o_1}} \cdot \frac{\partial In_{o_1}}{\partial h_1} \cdot \frac{\partial h_1}{\partial In_{h_1}} \cdot \frac{\partial In_{h_1}}{\partial w_1} + \frac{\partial Error}{\partial o_2} \cdot \frac{\partial o_2}{\partial In_{o_2}} \cdot \frac{\partial In_{o_2}}{\partial h_1} \cdot \frac{\partial h_1}{\partial In_{h_1}} \cdot \frac{\partial In_{h_1}}{\partial w_1}$$

$$= \left(\frac{\partial Error}{\partial o_1} \cdot \frac{\partial o_1}{\partial In_{o_1}} \cdot \frac{\partial In_{o_1}}{\partial h_1} + \frac{\partial Error}{\partial o_2} \cdot \frac{\partial o_2}{\partial In_{o_2}} \cdot \frac{\partial In_{o_2}}{\partial h_1}\right) \cdot \frac{\partial h_1}{\partial In_{h_1}} \cdot \frac{\partial In_{h_1}}{\partial w_1}$$

$$= [(o_1 - y_1) \times \text{sigmoid}'(In_{o1}) \times w_5 + (o_2 - y_2) \times \text{sigmoid}'(In_{o2}) \times w_6] \times \text{sigmoid}'(In_{h1}) \times x_1$$

NN training – backward propagation

- 更新權重

$$\boldsymbol{W}^{new} = \boldsymbol{W}^{old} - \eta \nabla_{\boldsymbol{W}} E$$

$\delta_1 \sim \delta_8$

$$\boldsymbol{b}^{new} = \boldsymbol{b}^{old} - \eta \nabla_{\boldsymbol{b}} E$$

Inner-product layer

- Forward propagation

$$X = \begin{bmatrix} x_1 & \dots & x_i \end{bmatrix} \quad W = \begin{bmatrix} w_{11} \dots w_{1j} \\ \vdots \quad \ddots \quad \vdots \\ w_{i1} \dots w_{ij} \end{bmatrix} \quad B = \begin{bmatrix} b_1 & \dots & b_j \end{bmatrix}$$

$$Y = XW + B$$

i : the number of input neurons
j : the number of output neurons

Inner-product layer

- Backward propagation

$$\nabla_{\boldsymbol{x}} E = \begin{bmatrix} \frac{\partial E}{\partial x_1} & \frac{\partial E}{\partial x_2} & \cdots & \frac{\partial E}{\partial x_i} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \ldots + \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial x_1}\right) & \cdots & \left(\frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \ldots + \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial x_i}\right) \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} & \cdots & \frac{\partial E}{\partial y_j} \end{bmatrix} \mathbf{A}$$

$$Y = XW + B$$

$$\frac{\partial E}{\partial X} = \frac{\partial E}{\partial Y}\frac{\partial Y}{\partial X} = \frac{\partial E}{\partial Y} W$$

$$\nabla_{\boldsymbol{w}} E = \begin{bmatrix} \frac{\partial E}{\partial w_{11}} & \cdots & \frac{\partial E}{\partial w_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial w_{i1}} & \cdots & \frac{\partial E}{\partial w_{ij}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial w_{11}} & \cdots & \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial w_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial w_{i1}} & \cdots & \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial w_{ij}} \end{bmatrix} = \mathbf{B} \begin{bmatrix} \frac{\partial E}{\partial y_1} & \cdots & \frac{\partial E}{\partial y_j} \end{bmatrix}$$

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial Y}\frac{\partial Y}{\partial W} = X\frac{\partial E}{\partial Y}$$

$$\nabla_{\boldsymbol{b}} E = \begin{bmatrix} \frac{\partial E}{\partial b_1} & \cdots & \frac{\partial E}{\partial b_j} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial b_1} & \cdots & \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial b_j} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} & \cdots & \frac{\partial E}{\partial y_j} \end{bmatrix} \mathbf{C}$$

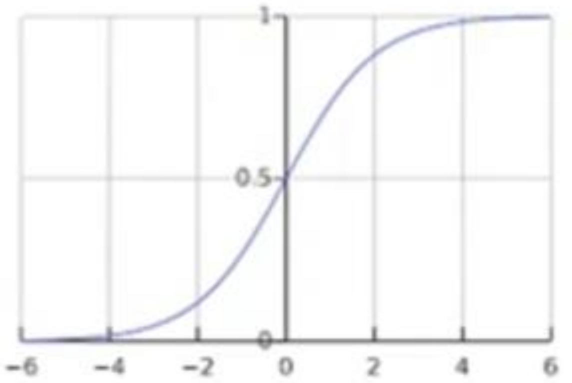
$$\frac{\partial E}{\partial B} = \frac{\partial E}{\partial Y}\frac{\partial Y}{\partial B} = \frac{\partial E}{\partial Y} 1$$

Please find the matrices **A**, **B** and **C**

Activation layer- Sigmoid function

- Forward propagation

$$\boldsymbol{y}(\boldsymbol{x}) = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1/(1+e^{-x_1}) \\ \vdots \\ 1/(1+e^{-x_n}) \end{bmatrix}$$



- Backward propagation

$$\nabla_{\boldsymbol{x}} E = \begin{bmatrix} \frac{\partial E}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1}\frac{\partial y_1}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial y_n}\frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \vdots \\ \frac{\partial E}{\partial y_n} \end{bmatrix} \boldsymbol{D}$$

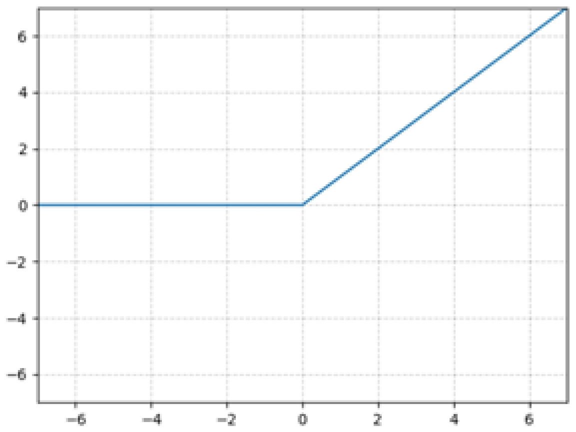
$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial x} = \frac{\partial E}{\partial y}\underbrace{\frac{d}{dx}f(x)}_{\boldsymbol{D}}$$

Please find the matrix $\boldsymbol{D}$

Activation layer- Rectified function

- Forward propagation

$$\boldsymbol{y}(\boldsymbol{x}) = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} [x_1 > 0] \cdot x_1 \\ \vdots \\ [x_n > 0] \cdot x_n \end{bmatrix}$$



- Backward propagation

$$\nabla_{\boldsymbol{x}} E = \begin{bmatrix} \frac{\partial E}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \frac{\partial y_1}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial y_n} \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \vdots \\ \frac{\partial E}{\partial y_n} \end{bmatrix} \boldsymbol{E}$$

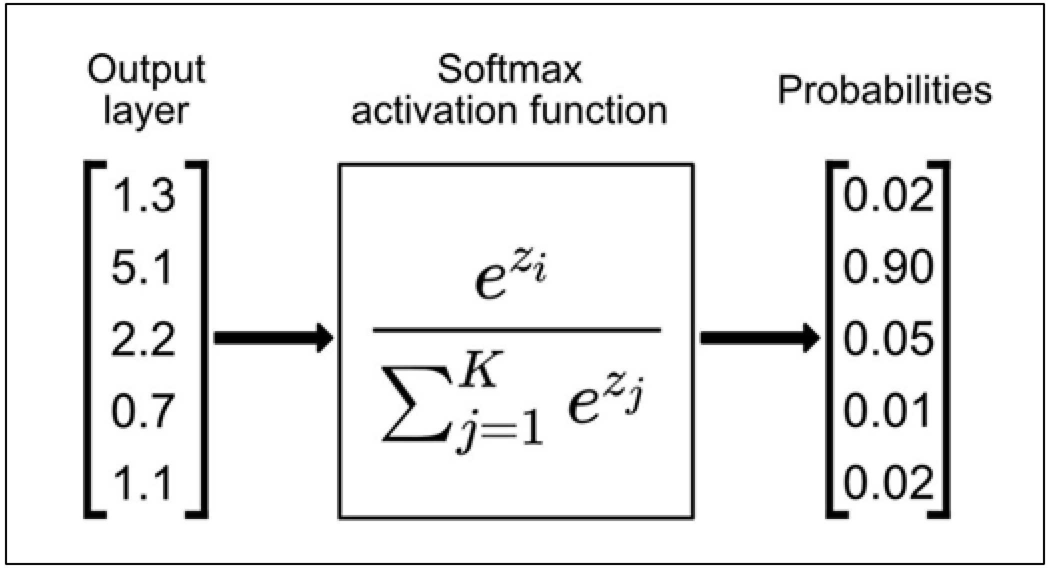
$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial x} = \frac{\partial E}{\partial y} \underbrace{\frac{d}{dx} f(x)}_{\boldsymbol{E}}$$

Please find the matrix $\boldsymbol{E}$

Softmax layer

- Forward propagation

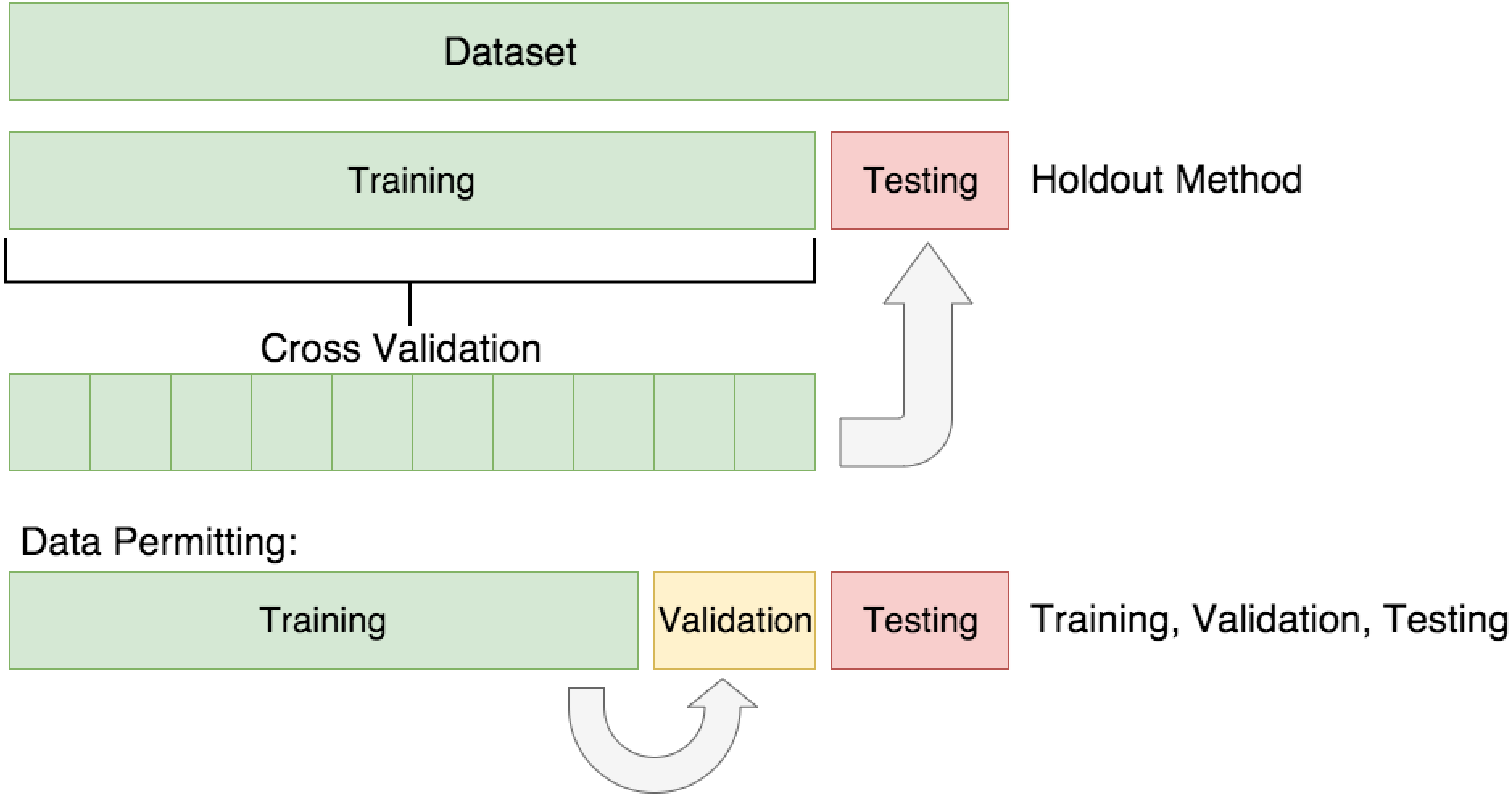
$$\boldsymbol{y}(\boldsymbol{x}) = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \frac{1}{\sum_{i=1}^{n} e^{x_i}} \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{bmatrix}$$



- Backward propagation

$$\nabla_{\boldsymbol{x}} E(\boldsymbol{y}, \boldsymbol{t}) = \begin{bmatrix} \frac{\partial E}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \frac{\partial y_1}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial y_n} \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} y_1 - t_1 \\ \vdots \\ y_n - t_n \end{bmatrix} = \boldsymbol{y} - \boldsymbol{t}$$

Cross validation



Outline

1. Lab1 tasks
2. Google Colab
3. How NN works
4. MNIST dataset

MNIST Dataset

- MNIST dataset is a large dataset of handwritten digits
- Train: 60000 images (28*28 pixels) + one label for each image
  - train_images[60000][784] : consist of 60000 images
  - train_labels[60000] : consist of 60000 labels
- Test: 10000 images (28*28 pixels) + one label for each image
  - test_images[10000][784] : consist of 10000 images
  - test_labels[10000] : consist of 10000 labels

原始 Dataset

- How to load ubyte?

t10k-images-idx3-ubyte	2017/9/24 下午 04:46	檔案	7,657 KB
t10k-labels-idx1-ubyte	2017/9/24 下午 04:46	檔案	10 KB
train-images-idx3-ubyte	2017/9/24 下午 04:46	檔案	45,938 KB
train-labels-idx1-ubyte	2017/9/24 下午 04:46	檔案	59 KB

TRAINING SET IMAGE FILE (train-images-idx3-ubyte):

[offset]	[type]	[value]	[description]
0000	32 bit integer	0x00000803(2051)	magic number
0004	32 bit integer	60000	number of images
0008	32 bit integer	28	number of rows
0012	32 bit integer	28	number of columns
0016	unsigned byte	??	pixel
0017	unsigned byte	??	pixel
……..			
xxxx	unsigned byte	??	pixel

```python
def convert(imgf, labelf, outf, n):
    f = open(imgf, "rb")
    o = open(outf, "w")
    l = open(labelf, "rb")

    f.read(16)
    l.read(8)
    images = []

    for i in range(n):
        image = [ord(l.read(1))]
        for j in range(28 * 28):
            image.append(ord(f.read(1)))
        images.append(image)

    for image in images:
        o.write(",".join(str(pix) for pix in image) + "\n")
    f.close()
    o.close()
    l.close()

convert("\MNIST\\train-images.idx3-ubyte", "\MNIST\\train-labels.idx1-ubyte",
        "\MNIST\mnist_train.csv", 60000)
convert("\MNIST\\t10k-images.idx3-ubyte", "\MNIST\\t10k-labels.idx1-ubyte",
        "\MNIST\mnist_test.csv", 10000)

print("Convert Finished!")
```

Coding Dataset

- mnist_train.csv & mnist_test.csv
- 讀寫容易，數據可以表格化展示

	A	B	C	D	E	F	G
1	7	0	0	0	0	0	0
2	2	0	0	0	0	0	0
3	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	4	0	0	0	0	0	0
6	1	0	0	0	0	0	0

pixel information

label

	MK	ML	MM	MN	MO	MP	MQ	MR	MS	MT
1	0	0	0	0	0	0	22	233	255	83
2	176	246	253	159	12	0	0	0	0	0
3	0	0	0	140	254	77	0	0	0	0
4	188	20	0	0	0	0	0	109	251	253
5	0	0	0	0	0	0	32	232	250	66
6	0	0	58	254	254	237	0	0	0	0

One-hot encoding

- easy to show predicted probability of each label

```
lr = np.arange(no_of_different_labels)

# transform labels into one hot representation
train_labels_one_hot = (lr==train_labels).astype(np.float)
val_labels_one_hot = (lr==val_labels).astype(np.float)
test_labels_one_hot = (lr==test_labels).astype(np.float)
```

```
label:  0  in one-hot representation:  [1 0 0 0 0 0 0 0 0 0]
label:  1  in one-hot representation:  [0 1 0 0 0 0 0 0 0 0]
label:  2  in one-hot representation:  [0 0 1 0 0 0 0 0 0 0]
label:  3  in one-hot representation:  [0 0 0 1 0 0 0 0 0 0]
label:  4  in one-hot representation:  [0 0 0 0 1 0 0 0 0 0]
label:  5  in one-hot representation:  [0 0 0 0 0 1 0 0 0 0]
label:  6  in one-hot representation:  [0 0 0 0 0 0 1 0 0 0]
label:  7  in one-hot representation:  [0 0 0 0 0 0 0 1 0 0]
label:  8  in one-hot representation:  [0 0 0 0 0 0 0 0 1 0]
label:  9  in one-hot representation:  [0 0 0 0 0 0 0 0 0 1]
```

Upload to moodle

- 學號_lab1.zip
  - 學號_lab1.ipynb
    - IPython notenook 須包含程式碼跟結果
  - 學號_lab1.pdf
    - 說明如何實作各個layer
    - 用自己的話說明forward / backward如何進行
    - 截圖並說明各項結果(包含accuracy和loss圖表(plot)的結果)
    - 實作過程中遇到的困難及你後來是如何解決的(optional)

Thanks for your listening

Advisor：Tsai, Chia-Chi

雲端連結:

https://drive.google.com/drive/folders/1Ptu1UW1lu8TUGzAyv3Xdd7ynhX2yV0Uo?usp=sharing