12주차(2/2)

심층 신경망 1

파이썬으로 배우는 기계학습

한동대학교
김영섭 교수

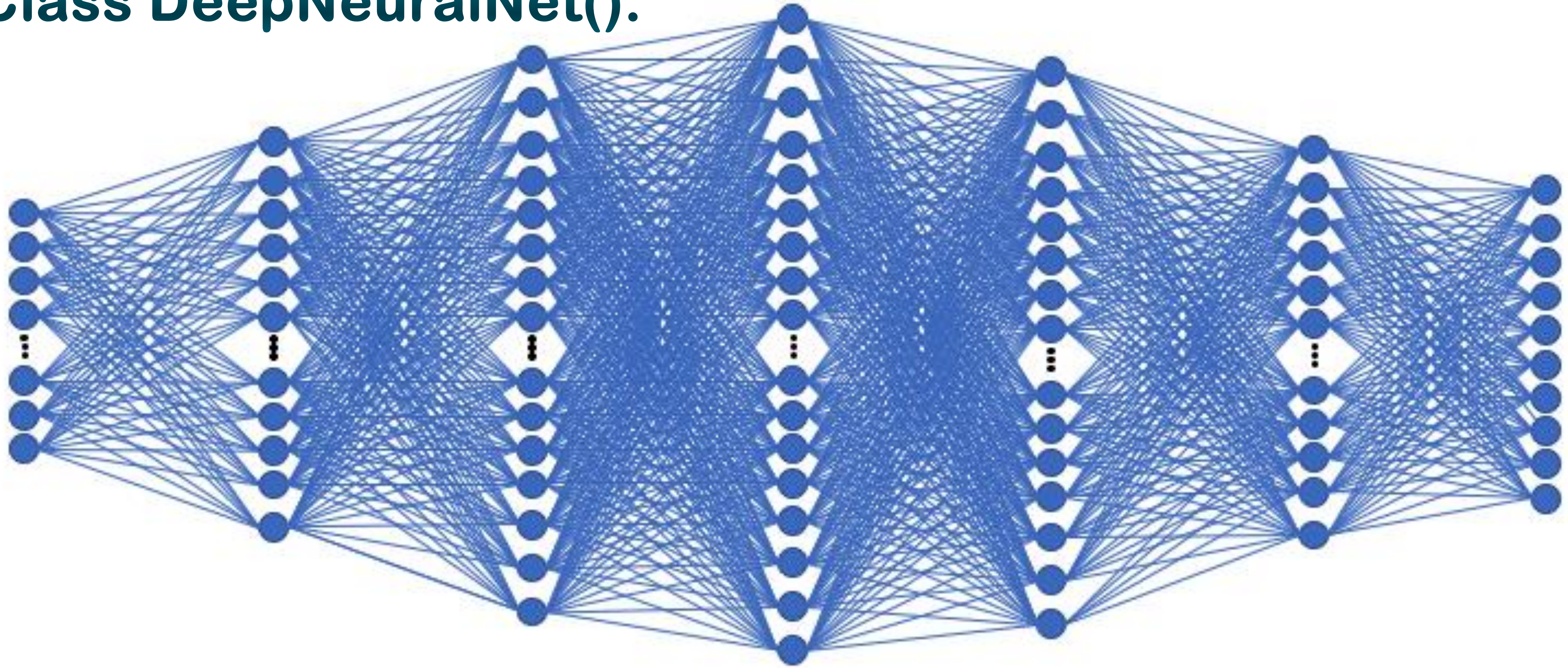
심층 신경망 1

- 학습 목표
  - 심층 신경망을 학습한다.
  - 은닉층의 개수에 따른 성능을 확인한다.

- 학습 내용
  - 심층 신경망 이해하기
  - 심층 신경망 구현하기
  - 심층 신경망 성능 확인하기

1. DNN 클래스

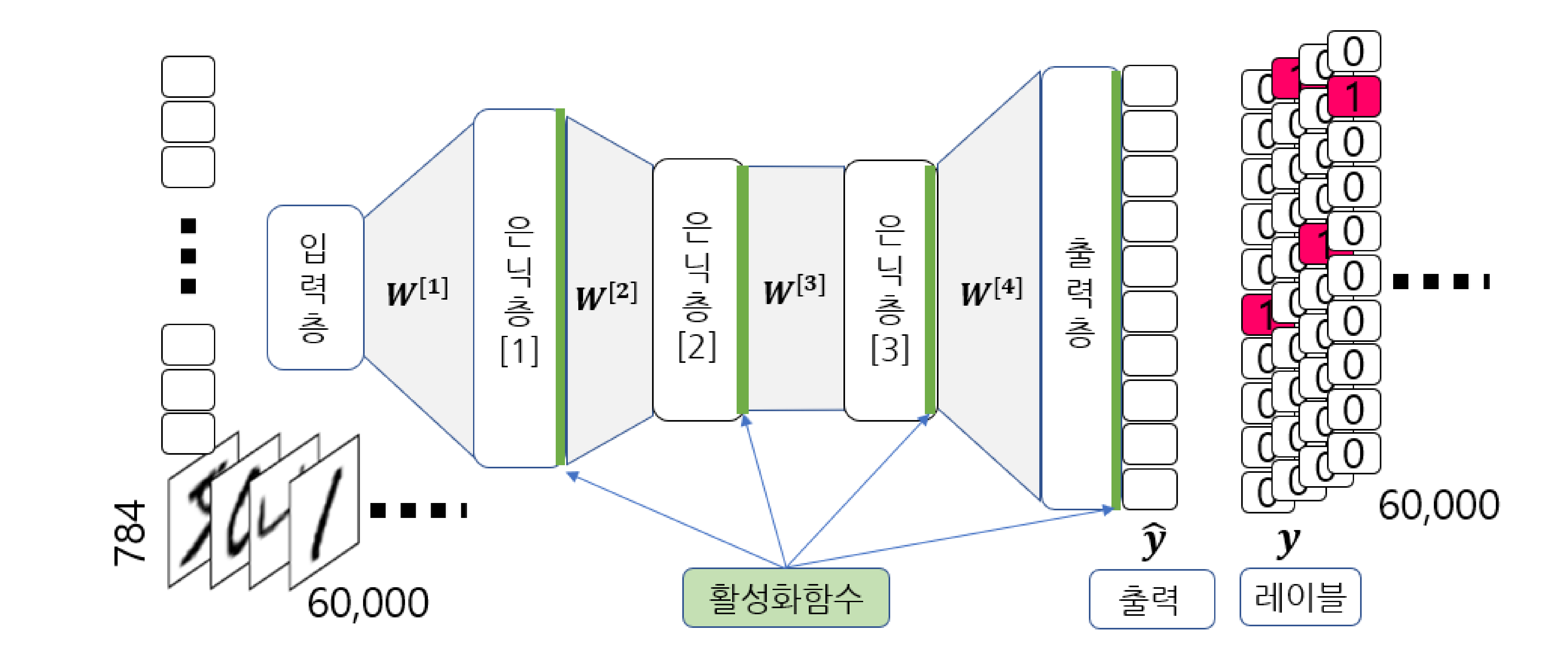
Class DeepNeuralNet():



1. DNN 클래스 : 코드 개요

```
class DeepNeuralNet():
    """ implements a deep neural net.
        Users may specify any number of layers.
        net_arch -- consists of a number of neurons in each layer
    """
    def __init__(self, net_arch, activate = None,
                     eta = 1.0, epochs = 100, random_seed = 1):
        pass

    def forpass(self, A0):
        pass

    def backprop(self, Z, A, Y):
        pass

    def fit(self, X, y):
        pass

    def predict(self, X):
        pass

    def evaluate(self, Xtest, ytest):
        pass
```

2. DNN 활성화 함수 : 구조



2. DNN 활성화 함수: tanh

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

```python
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수: tanh

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

$$\frac{d}{dx}\tanh(x) = 1 - \left(\frac{2}{1 + e^{-2x}} - 1\right)^{2}$$
$$= 1 - \tanh^{2}(x)$$

```
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수: sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{1}{1+\frac{1}{e^{x}}}$$

```python
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수: sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{1}{1+\frac{1}{e^{x}}}$$

$$\frac{d}{dx}\sigma(x) = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x))$$

```python
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
             1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수: ReLU

$$Relu(x) = \begin{cases} x & if\ x \geq 0 \\ 0 & otherwise \end{cases}$$

```
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수: ReLU

$$Relu(x) = \begin{cases} x & if\ x \geq 0 \\ 0 & otherwise \end{cases}$$

$$\frac{d}{dx} Relu(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & otherwise \end{cases}$$

```python
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

2. DNN 활성화 함수

```python
def tanh(x):
    return (1.0 - np.exp(-2 * x))/(
            1.0 + np.exp(-2 * x))

def tanh_d(x):
    return (1 + tanh(x)) * (1 - tanh(x))

def sigmoid(x):
    #x = np.clip(x, -500, 500)
    return 1 / (1 + np.exp((-x)))

def sigmoid_d(x):
    return sigmoid(x) * (1 - sigmoid(x))

def relu(x):
    return np.maximum(x, 0)

def relu_d(x):
    x[x<=0] = 0
    x[x>0] = 1
    return x
```

3. DNN 구현: 생성자

```python
def __init__(self, net_arch, activate = None,
             eta = 1.0, epochs = 100, random_seed = 1):
    self.eta = eta
    self.epochs = epochs
    self.net_arch = net_arch
    self.layers = len(net_arch)
    self.W = []

    self.g       = [lambda x: sigmoid(x)   for _ in range(self.layers)]
    self.g_prime = [lambda x: sigmoid_d(x) for _ in range(self.layers)]

    if activate is not None:
        for i, (g, g_prime) in enumerate(zip(activate[::2], activate[1::2])):
            self.g[i+1] = g
            self.g_prime[i+1] = g_prime

    np.random.seed(random_seed)
    self.W = [[None]]    ## the first W0 is not used.
    for layer in range(self.layers - 1):
        w = 2 * np.random.rand(self.net_arch[layer+1],
                               self.net_arch[layer]) - 1
        self.W.append(w)
```

3. DNN 구현: 생성자

```python
def __init__(self, net_arch, activate = None,
             eta = 1.0, epochs = 100, random_seed = 1):
    self.eta = eta
    self.epochs = epochs
    self.net_arch = net_arch
    self.layers = len(net_arch)
    self.W = []

    self.g       = [lambda x: sigmoid(x)   for _ in range(self.layers)]
    self.g_prime = [lambda x: sigmoid_d(x) for _ in range(self.layers)]

    if activate is not None:
        for i, (g, g_prime) in enumerate(zip(activate[::2], activate[1::2])):
            self.g[i+1] = g
            self.g_prime[i+1] = g_prime

    np.random.seed(random_seed)
    self.W = [[None]]    ## the first W0 is not used.
    for layer in range(self.layers - 1):
        w = 2 * np.random.rand(self.net_arch[layer+1],
                               self.net_arch[layer]) - 1
        self.W.append(w)
```

(뒷층 노드 수, 앞층 노드 수)

3. DNN 구현: fit() 메소드

```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(
            np.sum(cost * cost)))
    return self
```

4. fit() 메소드 : 순전파

```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(
            np.sum(cost * cost)))
    return self
```

4. fit() 메소드 : 순전파

```python
def forpass(self, A0):
    Z = [[None]] # Z0 is not used.
    A = []       # A0 = X0 is used.
    A.append(A0)
    for i in range(1, len(self.W)):
        z = np.dot(self.W[i], A[i-1])
        Z.append(z)
        a = self.g[i](z)
        A.append(a)
    return Z, A
```

```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(
            np.sum(cost * cost)))
    return self
```

$X \longrightarrow A^{[0]}$

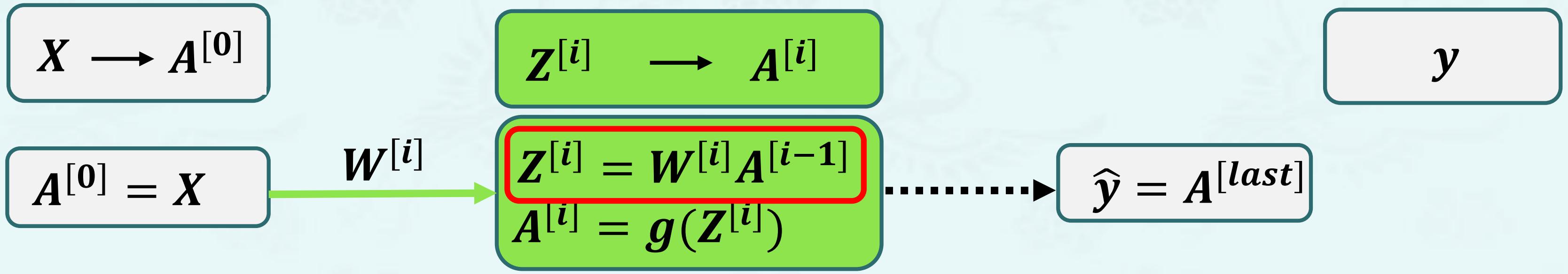
$Z^{[i]} \longrightarrow A^{[i]}$

y

$$A^{[0]} = X \xrightarrow{W^{[i]}} \begin{array}{l} Z^{[i]} = W^{[i]} A^{[i-1]} \\ A^{[i]} = g(Z^{[i]}) \end{array} \longrightarrow \hat{y} = A^{[last]}$$

4. fit() 메소드 : 순전파

```python
def forpass(self, A0):
    Z = [[None]] # Z0 is not used.
    A = []       # A0 = X0 is used.
    A.append(A0)
    for i in range(1, len(self.W)):
        z = np.dot(self.W[i], A[i-1])
        Z.append(z)
        a = self.g[i](z)
        A.append(a)
    return Z, A
```



4. fit() 메소드 : 순전파

```python
def forpass(self, A0):
    Z = [[None]] # Z0 is not used.
    A = []       # A0 = X0 is used.
    A.append(A0)
    for i in range(1, len(self.W)):
        z = np.dot(self.W[i], A[i-1])
        Z.append(z)
        a = self.g[i](z)
        A.append(a)
    return Z, A
```

$X \longrightarrow A^{[0]}$

$Z^{[i]} \longrightarrow A^{[i]}$

y

$A^{[0]} = X$ $\xrightarrow{W^{[i]}}$ $Z^{[i]} = W^{[i]} A^{[i-1]}$, $\boxed{A^{[i]} = g(Z^{[i]})}$ $\cdots\cdots\rightarrow$ $\hat{y} = A^{[last]}$

4. fit() 메소드 : 역전파

```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(
            np.sum(cost * cost)))
    return self
```

4. fit() 메소드 : 역전파

```python
def backprop(self, Z, A, Y):
    E  = [None for x in range(self.layers)]
    dZ = [None for x in range(self.layers)]

    ll = self.layers - 1
    error = Y - A[ll]
    E[ll] = error
    dZ[ll] = error * self.g_prime[ll](Z[ll])

    for i in range(self.layers-2, 0, -1):
        E[i]  = np.dot(self.W[i+1].T, E[i+1])
        dZ[i] = E[i] * self.g_prime[i](Z[i])

    m = Y.shape[0] # number of samples
    for i in range(ll, 0, -1):
        self.W[i] += self.eta * np.dot(dZ[i], A[i-1].T) / m
    return error
```

```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(np.sum(cost * cost)))
    return self
```

$$X \longrightarrow A^{[0]}$$

$$A^{[0]} = X$$

$$W^{[i]}$$

$$Z^{[i]} = W^{[i]}A^{[i-1]}$$
$$A^{[i]} = g(Z^{[i]})$$

$$\hat{y} = A^{[last]}$$

$$y$$

$$W^{[last]} += dZ^{[last]} \cdot A^{[last-1]T}$$

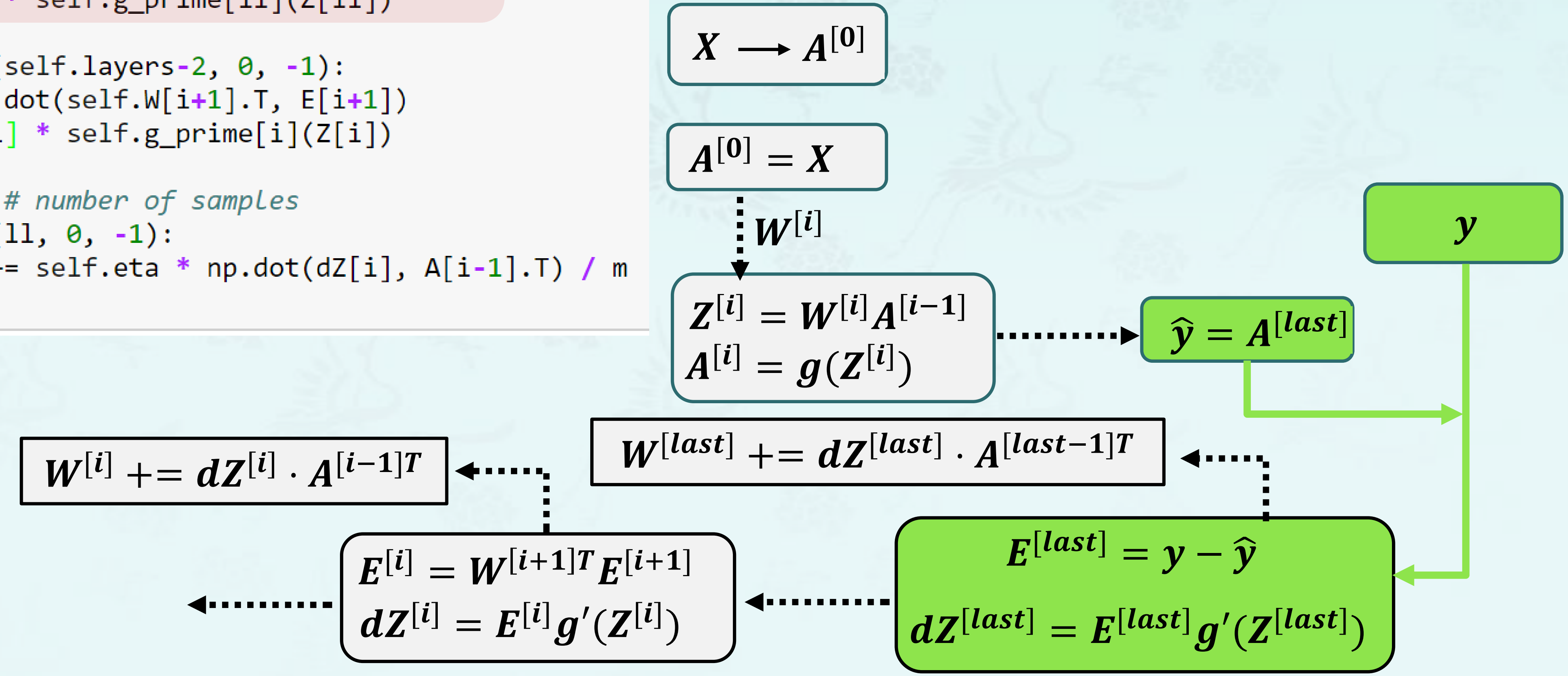
$$E^{[last]} = y - \hat{y}$$
$$dZ^{[last]} = E^{[last]} g'(Z^{[last]})$$

$$E^{[i]} = W^{[i+1]T} E^{[i+1]}$$
$$dZ^{[i]} = E^{[i]} g'(Z^{[i]})$$

$$W^{[i]} += dZ^{[i]} \cdot A^{[i-1]T}$$

4. fit() 메소드 : 역전파

```python
def backprop(self, Z, A, Y):
    E  = [None for x in range(self.layers)]
    dZ = [None for x in range(self.layers)]

    ll = self.layers - 1
    error = Y - A[ll]
    E[ll] = error
    dZ[ll] = error * self.g_prime[ll](Z[ll])

    for i in range(self.layers-2, 0, -1):
        E[i]  = np.dot(self.W[i+1].T, E[i+1])
        dZ[i] = E[i] * self.g_prime[i](Z[i])

    m = Y.shape[0] # number of samples
    for i in range(ll, 0, -1):
        self.W[i] += self.eta * np.dot(dZ[i], A[i-1].T) / m
    return error
```



4. fit() 메소드 : 역전파

```python
def backprop(self, Z, A, Y):
    E  = [None for x in range(self.layers)]
    dZ = [None for x in range(self.layers)]

    ll = self.layers - 1
    error = Y - A[ll]
    E[ll] = error
    dZ[ll] = error * self.g_prime[ll](Z[ll])

    for i in range(self.layers-2, 0, -1):
        E[i]  = np.dot(self.W[i+1].T, E[i+1])
        dZ[i] = E[i] * self.g_prime[i](Z[i])

    m = Y.shape[0] # number of samples
    for i in range(ll, 0, -1):
        self.W[i] += self.eta * np.dot(dZ[i], A[i-1].T) / m
    return error
```

$X \longrightarrow A^{[0]}$

$A^{[0]} = X$

$W^{[i]}$

$Z^{[i]} = W^{[i]} A^{[i-1]}$

$A^{[i]} = g(Z^{[i]})$

$\hat{y} = A^{[last]}$

y

$W^{[last]} \mathrel{+}= dZ^{[last]} \cdot A^{[last-1]T}$

$E^{[last]} = y - \hat{y}$

$dZ^{[last]} = E^{[last]} g'(Z^{[last]})$

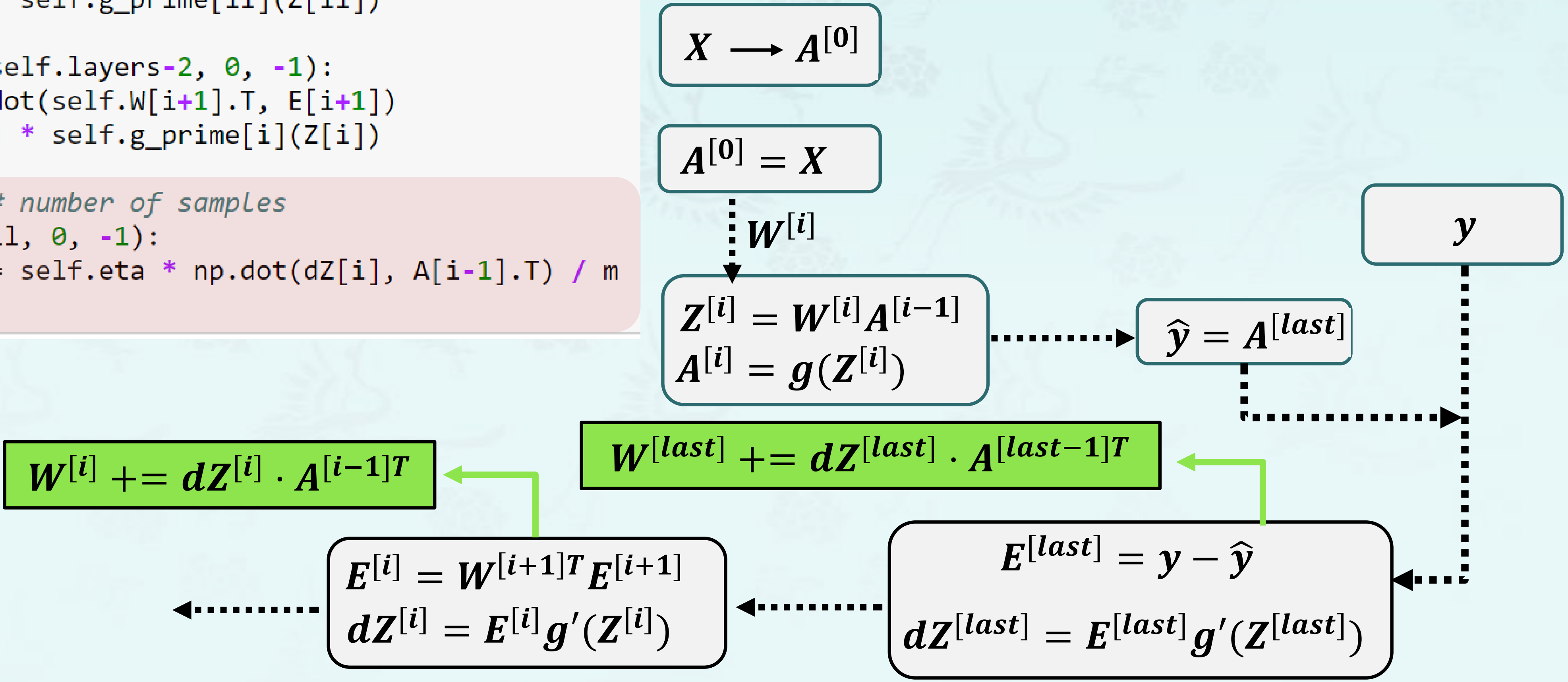
$W^{[i]} \mathrel{+}= dZ^{[i]} \cdot A^{[i-1]T}$

$E^{[i]} = W^{[i+1]T} E^{[i+1]}$

$dZ^{[i]} = E^{[i]} g'(Z^{[i]})$

4. fit() 메소드 : 역전파

```python
def backprop(self, Z, A, Y):
    E  = [None for x in range(self.layers)]
    dZ = [None for x in range(self.layers)]

    ll = self.layers - 1
    error = Y - A[ll]
    E[ll] = error
    dZ[ll] = error * self.g_prime[ll](Z[ll])

    for i in range(self.layers-2, 0, -1):
        E[i]  = np.dot(self.W[i+1].T, E[i+1])
        dZ[i] = E[i] * self.g_prime[i](Z[i])

    m = Y.shape[0] # number of samples
    for i in range(ll, 0, -1):
        self.W[i] += self.eta * np.dot(dZ[i], A[i-1].T) / m
    return error
```



4. fit() 메소드 : 오차 계산

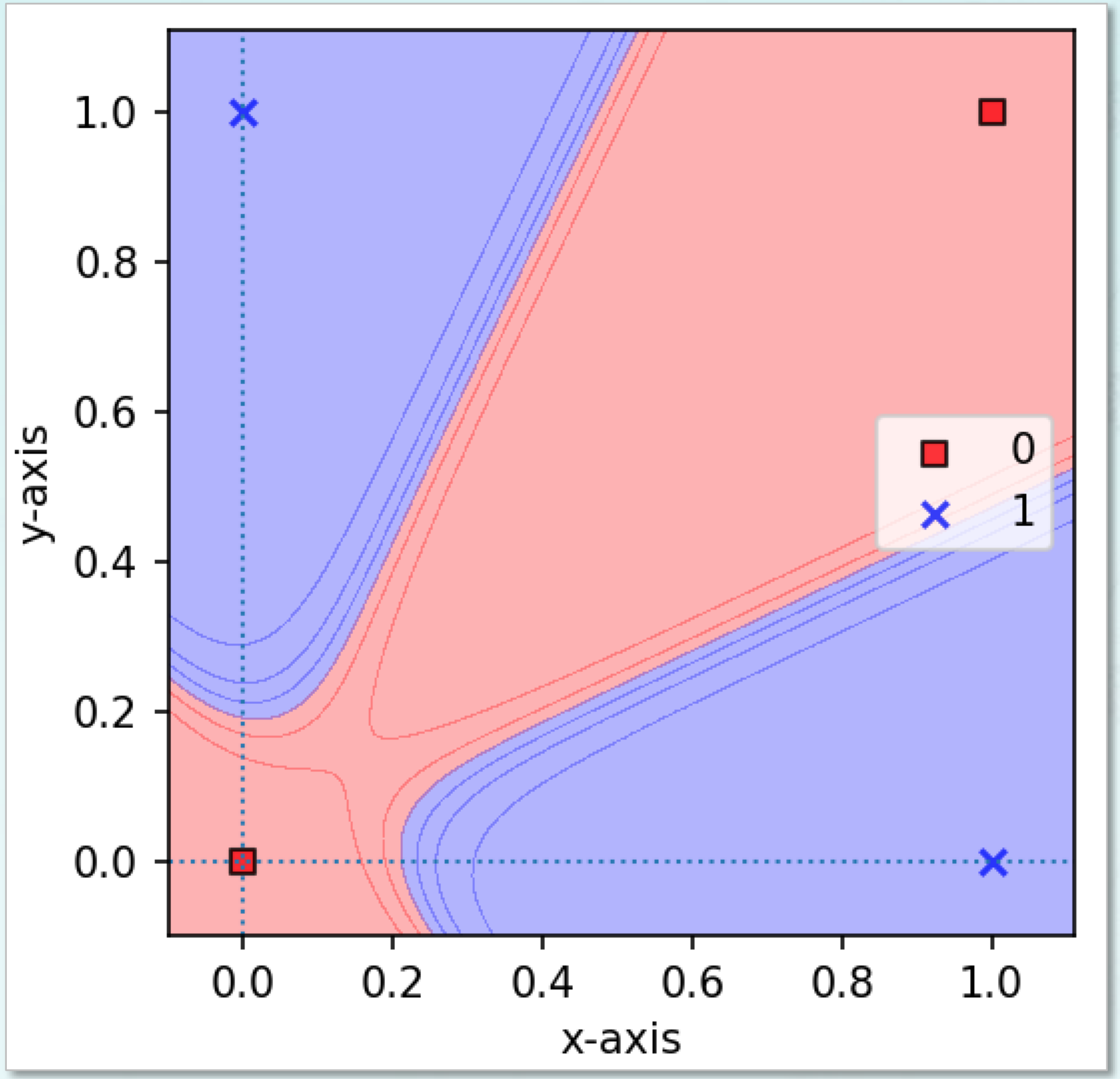
```python
def fit(self, X, y):
    self.cost_ = []
    for epoch in range(self.epochs):
        Z, A = self.forpass(X)
        cost = self.backprop(Z, A, y)
        self.cost_.append(np.sqrt(
              np.sum(cost * cost)))
    return self
```

5. DNN 학습 결과: XOR 테스트

```python
import joy
X = np.array([ [0, 0, 1, 1], [0, 1, 0, 1] ])
y = np.array([0, 1, 1, 0])
dnn = DeepNeuralNet([2, 4, 2, 1], eta = 0.5, epochs = 5000).fit(X, y)

joy.plot_decision_regions(X.T, y, dnn)
plt.xlabel('x-axis')
plt.ylabel('y-axis')
plt.legend(loc='best')
plt.show()
```

[입력층, 은닉층, 은닉층, 출력층]의 노드 수

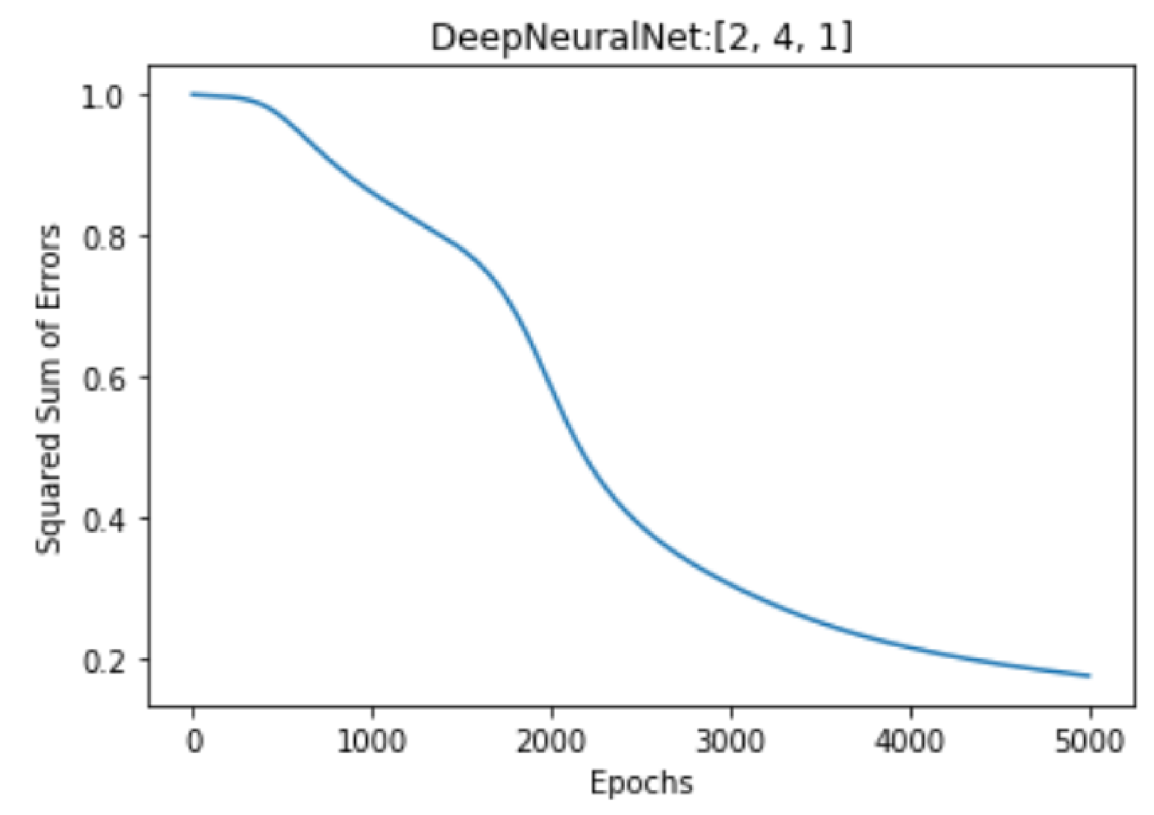
5. DNN 학습 결과: XOR 테스트



5. DNN 학습 결과: XOR 테스트

```
dnn = DeepNeuralNet([2, 4, 1],
                    eta = 0.5, epochs = 5000).fit(X, y)
plt.plot(range(len(dnn.cost_)), dnn.cost_)
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.title('DeepNeuralNet:{}'.format(dnn.net_arch))
plt.show()
```

은닉층 1개



5. DNN 학습 결과: 비교를 위한 학습 코딩

```
dnn1 = DeepNeuralNet([2,4,1], eta = 0.5, epochs = 5000).fit(X, y)
g = [sigmoid, sigmoid_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn2 = DeepNeuralNet([2,4,2,1], activate=g, eta = 0.5, epochs = 5000).fit(X, y)
plt.plot(range(len(dnn1.cost_)), dnn1.cost_, label='{}'.format(dnn1.net_arch))
plt.plot(range(len(dnn2.cost_)), dnn2.cost_, label='{}'.format(dnn2.net_arch))
plt.title('DeepNeuralNet:{} vs {}'.format(dnn1.net_arch, dnn2.net_arch))
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.legend(loc='best')
plt.show()
```

5. DNN 학습 결과: 비교를 위한 학습 코딩

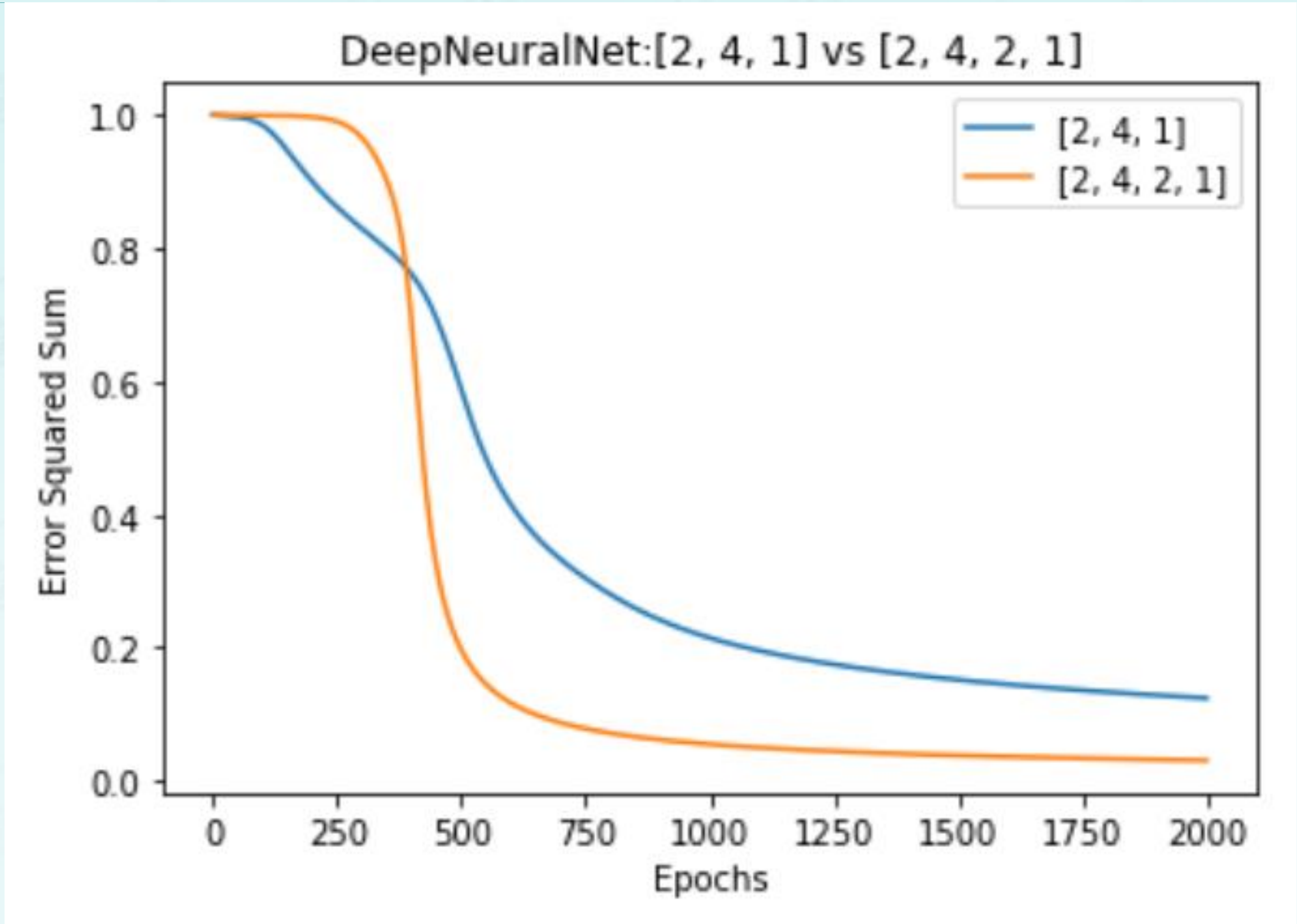
```
dnn1 = DeepNeuralNet([2,4,1], eta = 0.5, epochs = 5000).fit(X, y)
g = [sigmoid, sigmoid_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn2 = DeepNeuralNet([2,4,2,1], activate=g, eta = 0.5, epochs = 5000).fit(X, y)
plt.plot(range(len(dnn1.cost_)), dnn1.cost_, label='{}'.format(dnn1.net_arch))
plt.plot(range(len(dnn2.cost_)), dnn2.cost_, label='{}'.format(dnn2.net_arch))
plt.title('DeepNeuralNet:{} vs {}'.format(dnn1.net_arch, dnn2.net_arch))
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.legend(loc='best')
plt.show()
```

은닉층 2개

5. DNN 학습 결과: 비교를 위한 학습 코딩

```
dnn1 = DeepNeuralNet([2,4,1], eta = 0.5, epochs = 5000).fit(X, y)
g = [sigmoid, sigmoid_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn2 = DeepNeuralNet([2,4,2,1], activate=g, eta = 0.5, epochs = 5000).fit(X, y)
plt.plot(range(len(dnn1.cost_)), dnn1.cost_, label='{}'.format(dnn1.net_arch))
plt.plot(range(len(dnn2.cost_)), dnn2.cost_, label='{}'.format(dnn2.net_arch))
plt.title('DeepNeuralNet:{} vs {}'.format(dnn1.net_arch, dnn2.net_arch))
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.legend(loc='best')
plt.show()
```

5. DNN 학습 결과: 시각화 결과

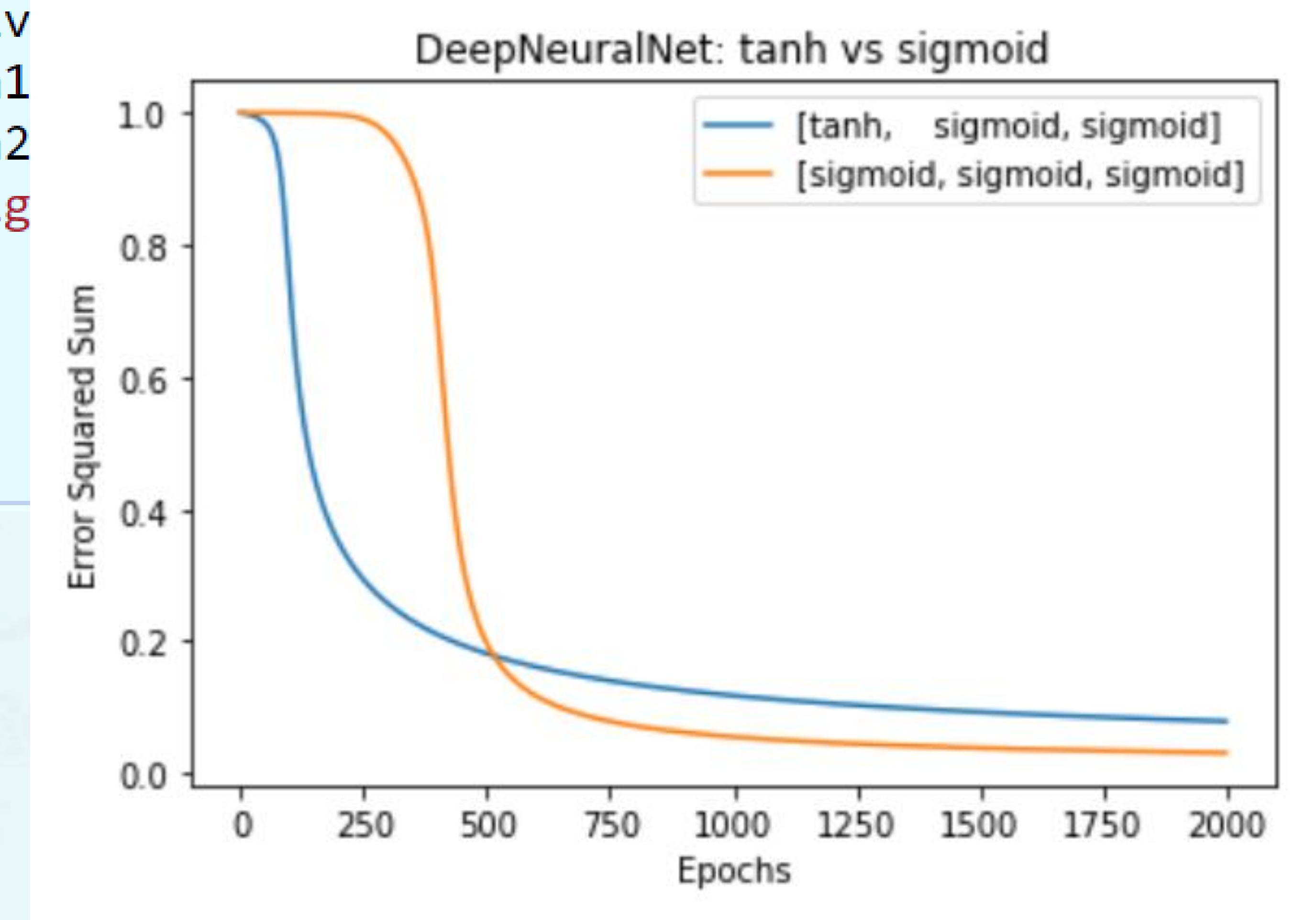


6. DNN 학습: 각 층별로 활성화 함수 지정

```
g1 = [tanh, tanh_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn1 = DeepNeuralNet([2,4,2,1], activate=g1, eta = 0.5, epochs = 5000).fit(X, y)
g2 = [sigmoid, sigmoid_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn2 = DeepNeuralNet([2,4,2,1], activate=g2, eta = 0.5, epochs = 5000).fit(X, y)
plt.plot(range(len(dnn1.cost_)), dnn1.cost_, label='[tanh,    sigmoid, sigmoid]')
plt.plot(range(len(dnn2.cost_)), dnn2.cost_, label='[sigmoid, sigmoid, sigmoid]')
plt.title('DeepNeuralNet: tanh vs sigmoid')
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.legend(loc='best')
plt.show()
```

6. DNN 학습: 각 층별로 활성화 함수 지정

```
g1 = [tanh, tanh_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn1 = DeepNeuralNet([2,4,2,1], activate=g1, eta = 0.5, epochs = 5000).fit(X, y)
g2 = [sigmoid, sigmoid_d, sigmoid, sigmoid_d, sigmoid, sigmoid_d]
dnn2 = DeepNeuralNet([2,4,2,1], activ
plt.plot(range(len(dnn1.cost_)), dnn1
plt.plot(range(len(dnn2.cost_)), dnn2
plt.title('DeepNeuralNet: tanh vs sig
plt.xlabel('Epochs')
plt.ylabel('Squared Sum of Errors')
plt.legend(loc='best')
plt.show()
```



심층 신경망 1

- 학습 정리
  - 심층 신경망인 **DeepNeuralNet** 클래스 구현
  - 심층 신경망의 은닉층 갯수에 따른 성능 확인