

---

# **DeepLearningPython35**

**Daniel Engbert, Michael Nielsen, Michal Dobrzanski**

**Sep 03, 2021**



**CONTENTS:**

<b>1</b>	<b>DeepLearningPython35</b>	<b>1</b>
1.1	bootyNet module . . . . .	1
1.2	expand_mnist module . . . . .	1
1.3	mnist_average_darkness module . . . . .	1
1.4	mnist_loader module . . . . .	2
1.5	mnist_svm module . . . . .	3
1.6	mynet module . . . . .	3
1.7	network module . . . . .	5
1.8	network2 module . . . . .	6
1.9	network3 module . . . . .	8
1.10	test module . . . . .	8
1.11	test1 module . . . . .	8
1.12	tests package . . . . .	8
<b>2</b>	<b>Indices and tables</b>	<b>11</b>
	<b>Python Module Index</b>	<b>13</b>
	<b>Index</b>	<b>15</b>



## DEEPLEARNINGPYTHON35

### 1.1 bootyNet module

```
bootyNet.main()  
bootyNet.gooo (prev=None, entropyYears=0)  
bootyNet.getConvergeRate (history, largeBin, smallBin, threshHold=0.005)  
bootyNet.visualizeResults (dataSet, fname='backups/latest.pkl', saveSuccesses=False)
```

### 1.2 expand\_mnist module

#### 1.2.1 expand\_mnist.py

Take the 50,000 MNIST training images, and create an expanded set of 250,000 images, by displacing each training image up, down, left and right, by one pixel. Save the resulting file to ../data/mnist\_expanded.pkl.gz.

Note that this program is memory intensive, and may not run on small systems.

```
expand_mnist.main()
```

### 1.3 mnist\_average\_darkness module

#### 1.3.1 mnist\_average\_darkness

A naive classifier for recognizing handwritten digits from the MNIST data set. The program classifies digits based on how dark they are — the idea is that digits like “1” tend to be less dark than digits like “8”, simply because the latter has a more complex shape. When shown an image the classifier returns whichever digit in the training data had the closest average darkness.

The program works in two steps: first it trains the classifier, and then it applies the classifier to the MNIST test data to see how many digits are correctly classified.

Needless to say, this isn’t a very good way of recognizing handwritten digits! Still, it’s useful to show what sort of performance we get from naive ideas.

```
mnist_average_darkness.main()
```

`mnist_average_darkness.avg_darknesses(training_data)`

Return a defaultdict whose keys are the digits 0 through 9. For each digit we compute a value which is the average darkness of training images containing that digit. The darkness for any particular image is just the sum of the darknesses for each pixel.

`mnist_average_darkness.guess_digit(image, avgs)`

Return the digit whose average darkness in the training data is closest to the darkness of `image`. Note that `avgs` is assumed to be a defaultdict whose keys are 0...9, and whose values are the corresponding average darknesses across the training data.

## 1.4 mnist\_loader module

### 1.4.1 mnist\_loader

A library to load the MNIST image data. For details of the data structures that are returned, see the doc strings for `load_data` and `load_data_wrapper`. In practice, `load_data_wrapper` is the function usually called by our neural network code.

`mnist_loader.load_data()`

Return the MNIST data as a tuple containing the training data, the validation data, and the test data. The `training_data` is returned as a tuple with two entries. The first entry contains the actual training images. This is a numpy ndarray with 50,000 entries. Each entry is, in turn, a numpy ndarray with 784 values, representing the  $28 * 28 = 784$  pixels in a single MNIST image. The second entry in the `training_data` tuple is a numpy ndarray containing 50,000 entries. Those entries are just the digit values (0...9) for the corresponding images contained in the first entry of the tuple. The `validation_data` and `test_data` are similar, except each contains only 10,000 images. This is a nice data format, but for use in neural networks it's helpful to modify the format of the `training_data` a little. That's done in the wrapper function `load_data_wrapper()`, see below.

`mnist_loader.load_data_wrapper()`

Return a tuple containing (`training_data`, `validation_data`, `test_data`). Based on `load_data`, but the format is more convenient for use in our implementation of neural networks. In particular, `training_data` is a list containing 50,000 2-tuples (`x`, `y`). `x` is a 784-dimensional numpy.ndarray containing the input image. `y` is a 10-dimensional numpy.ndarray representing the unit vector corresponding to the correct digit for `x`. `validation_data` and `test_data` are lists containing 10,000 2-tuples (`x`, `y`). In each case, `x` is a 784-dimensional numpy.ndarry containing the input image, and `y` is the corresponding classification, i.e., the digit values (integers) corresponding to `x`. Obviously, this means we're using slightly different formats for the training data and the validation / test data. These formats turn out to be the most convenient for use in our neural network code.

`mnist_loader.vectorized_result(j)`

Return a 10-dimensional unit vector with a 1.0 in the `j`th position and zeroes elsewhere. This is used to convert a digit (0...9) into a corresponding desired output from the neural network.

`mnist_loader.vectorToImage(v, fname="")`

## 1.5 mnist\_svm module

### 1.5.1 mnist\_svm

A classifier program for recognizing handwritten digits from the MNIST data set, using an SVM classifier.

```
mnist_svm.svm_baseline()
```

## 1.6 mynet module

### 1.6.1 mynet.py

A module to implement the stochastic gradient descent learning algorithm for a feedforward neural network.

More info: <http://neuralnetworksanddeeplearning.com/chap1.html> <http://neuralnetworksanddeeplearning.com/chap2.html>

```
class mynet.Network(sizes)
```

Bases: object

```
__init__(sizes)
```

The list `sizes` contains the number of neurons in the respective layers of the network. For example, if the list was `[2, 3, 1]` then it would be a three-layer network, with the first layer containing 2 neurons, the second layer 3 neurons, and the third layer 1 neuron. The biases and weights for the network are initialized randomly, using a Gaussian distribution with mean 0, and variance 1. Note that the first layer is assumed to be an input layer, and by convention we won't set any biases for those neurons, since biases are only ever used in computing the outputs from later layers.

**Parameters** `sizes` (list of int) – list of layer sizes (e.g. `[2,3,2]`)

**sizes**

list of layer sizes (e.g. `[2,3,2]`)

**Type** list of int

**num\_layers**

number of layers in network

**Type** int

**biases**

stores a column vector for each layer's biases (except the input layer). Initialized randomly using `np.random.randn` (for a gaussian distribution) example for sizes `[2,3,2]`:

$$\left[ \begin{bmatrix} -0.25 \\ -0.34 \\ 0.21 \end{bmatrix}, \begin{bmatrix} -0.88 \\ 0.47 \end{bmatrix} \right]$$

**Type** list of `numpy.ndarray`

**weights**

list of weights in the network (one np array for each layer)

for a given layer, col 0 = vector of weights for connections from node 0 (in cur layer) to next layer's nodes ample for sizes `[2,3,2]`:

$$\left[ \begin{bmatrix} -0.3, -0.01 \\ 2.94, -0.88 \\ 0.91, -1.85 \end{bmatrix}, \begin{bmatrix} 1.62, 1.21, 0.95 \\ -0.19, -0.68, 0.61 \end{bmatrix} \right]$$

let  $w = weights[0]$  (stores all weights coming into layer 1 from layer 0).

$w[j]$  is the list of weights coming into node  $j$  in layer 1, so  $w_{jk} = w[j][k]$  is the weight between the  $j$ th neuron in layer 1, and the  $k$ th neuron in layer 0.

This ordering allows us to calculate activations with:  $a' = \sigma(wa + b)$

**Type** list of `numpy.ndarray`

**save** (*filename*)

pickle data in this class as a backup to desired filename <https://stackoverflow.com/a/2842727>

**load** (*filename*)

load data from pickle file into this class to overwrite its data Incomplete: (returns a Network object)

**test** (*test\_data*)

Return the number of test inputs for which the neural network outputs the correct result. Note that the neural network's output is assumed to be the index of whichever neuron in the final layer has the highest activation.

**getOutput** (*x*)

returns the output layers activations for given input (using feedforward())

**feedforward** (*a*)

returns list of z values and list of activations at each layer for given input

**updateMiniBatch** (*miniBatch*, *rate*)

Perform gradient descent using backpropagation on a single miniBatch and update the network's weights and biases.

#### Parameters

- **miniBatch** (list of tuples) – List of training inputs / expected outputs (x,y). Calculated changes to the weights and biases will be averaged across this mini batch.

Note: to train on a single sample, set `mini_batch=[(x,y)]`

- **rate** (float) – Learning rate to use for training (e.g. 0.05).

**SGD** (*training\_data*, *epochs*, *mini\_batch\_size*, *rate*, *test\_data=None*, *start\_epoch=1*)

Train the neural network using mini-batch stochastic gradient descent. The `training_data` is a list of tuples (*x*, *y*) representing the training inputs and the desired outputs. The other non-optional parameters are self-explanatory. If `test_data` is provided then the network will be evaluated against the test data after each epoch, and partial progress printed out. This is useful for tracking progress, but slows things down substantially.

**backprop** (*x*, *y*)

do backpropagation returns `nabla_b`, `nabla_w` (partial derivatives of biases and weights wrt. cost function)

**cost\_derivative** (*output\_activations*, *y*)

Return the vector of partial derivatives  $\partial C_x / \partial a$  for the output activations. (For the quadratic cost function  $C$  in particular)

#### Parameters

- **output\_activations** – vector of output activations
- **y** – expected output vector

**sigmoid** (*z*)

Implements the sigmoid function. [https://en.wikipedia.org/wiki/Sigmoid\\_function](https://en.wikipedia.org/wiki/Sigmoid_function)

**sigmoid\_prime** (*z*)

Derivative of the sigmoid function.



## 1.7 network module

### 1.7.1 network.py

#### IT WORKS

A module to implement the stochastic gradient descent learning algorithm for a feedforward neural network. Gradients are calculated using backpropagation. Note that I have focused on making the code simple, easily readable, and easily modifiable. It is not optimized, and omits many desirable features.

**class** `network.Network` (*sizes*)

Bases: `object`

**\_\_init\_\_** (*sizes*)

The list *sizes* contains the number of neurons in the respective layers of the network. For example, if the list was [2, 3, 1] then it would be a three-layer network, with the first layer containing 2 neurons, the second layer 3 neurons, and the third layer 1 neuron. The biases and weights for the network are initialized randomly, using a Gaussian distribution with mean 0, and variance 1. Note that the first layer is assumed to be an input layer, and by convention we won't set any biases for those neurons, since biases are only ever used in computing the outputs from later layers.

**save** (*filename*)

**load** (*filename*)

**feedforward** (*a*)

Return the output of the network if *a* is input.

**SGD** (*training\_data*, *epochs*, *mini\_batch\_size*, *eta*, *test\_data=None*, *start\_epoch=1*)

Train the neural network using mini-batch stochastic gradient descent. The *training\_data* is a list of tuples (*x*, *y*) representing the training inputs and the desired outputs. The other non-optional parameters are self-explanatory. If *test\_data* is provided then the network will be evaluated against the test data after each epoch, and partial progress printed out. This is useful for tracking progress, but slows things down substantially.

**update\_mini\_batch** (*mini\_batch*, *eta*)

Update the network's weights and biases by applying gradient descent using backpropagation to a single mini batch. The *mini\_batch* is a list of tuples (*x*, *y*), and *eta* is the learning rate.

**backprop** (*x*, *y*)

Return a tuple (*nabla\_b*, *nabla\_w*) representing the gradient for the cost function *C<sub>x</sub>*. *nabla\_b* and *nabla\_w* are layer-by-layer lists of numpy arrays, similar to *self.biases* and *self.weights*.

**evaluate** (*test\_data*)

Return the number of test inputs for which the neural network outputs the correct result. Note that the neural network's output is assumed to be the index of whichever neuron in the final layer has the highest activation.

**cost\_derivative** (*output\_activations*, *y*)

Return the vector of partial derivatives partial *C<sub>x</sub>* / partial *a* for the output activations.

`network.sigmoid` (*z*)

The sigmoid function.

`network.sigmoid_prime` (*z*)

Derivative of the sigmoid function.

## 1.8 network2 module

### 1.8.1 network2.py

An improved version of `network.py`, implementing the stochastic gradient descent learning algorithm for a feedforward neural network. Improvements include the addition of the cross-entropy cost function, regularization, and better initialization of network weights. Note that I have focused on making the code simple, easily readable, and easily modifiable. It is not optimized, and omits many desirable features.

```
class network2.QuadraticCost
```

```
    Bases: object
```

```
    static fn (a, y)
```

```
        Return the cost associated with an output a and desired output y.
```

```
    static delta (z, a, y)
```

```
        Return the error delta from the output layer.
```

```
class network2.CrossEntropyCost
```

```
    Bases: object
```

```
    static fn (a, y)
```

```
        Return the cost associated with an output a and desired output y. Note that np.nan_to_num is used to ensure numerical stability. In particular, if both a and y have a 1.0 in the same slot, then the expression (1-y)*np.log(1-a) returns nan. The np.nan_to_num ensures that that is converted to the correct value (0.0).
```

```
    static delta (z, a, y)
```

```
        Return the error delta from the output layer. Note that the parameter z is not used by the method. It is included in the method's parameters in order to make the interface consistent with the delta method for other cost classes.
```

```
class network2.Network (sizes, cost=<class 'network2.CrossEntropyCost'>)
```

```
    Bases: object
```

```
    __init__ (sizes, cost=<class 'network2.CrossEntropyCost'>)
```

```
        The list sizes contains the number of neurons in the respective layers of the network. For example, if the list was [2, 3, 1] then it would be a three-layer network, with the first layer containing 2 neurons, the second layer 3 neurons, and the third layer 1 neuron. The biases and weights for the network are initialized randomly, using self.default_weight_initializer (see docstring for that method).
```

```
    default_weight_initializer ()
```

```
        Initialize each weight using a Gaussian distribution with mean 0 and standard deviation 1 over the square root of the number of weights connecting to the same neuron. Initialize the biases using a Gaussian distribution with mean 0 and standard deviation 1.
```

```
        Note that the first layer is assumed to be an input layer, and by convention we won't set any biases for those neurons, since biases are only ever used in computing the outputs from later layers.
```

```
    large_weight_initializer ()
```

```
        Initialize the weights using a Gaussian distribution with mean 0 and standard deviation 1. Initialize the biases using a Gaussian distribution with mean 0 and standard deviation 1.
```

```
        Note that the first layer is assumed to be an input layer, and by convention we won't set any biases for those neurons, since biases are only ever used in computing the outputs from later layers.
```

```
        This weight and bias initializer uses the same approach as in Chapter 1, and is included for purposes of comparison. It will usually be better to use the default weight initializer instead.
```

```
    feedforward (a)
```

```
        Return the output of the network if a is input.
```

**SGD** (*training\_data*, *epochs*, *mini\_batch\_size*, *eta*, *lmbda*=0.0, *evaluation\_data*=None, *monitor\_evaluation\_cost*=False, *monitor\_evaluation\_accuracy*=False, *monitor\_training\_cost*=False, *monitor\_training\_accuracy*=False, *early\_stopping\_n*=0)

Train the neural network using mini-batch stochastic gradient descent. The *training\_data* is a list of tuples (*x*, *y*) representing the training inputs and the desired outputs. The other non-optional parameters are self-explanatory, as is the regularization parameter *lmbda*. The method also accepts *evaluation\_data*, usually either the validation or test data. We can monitor the cost and accuracy on either the evaluation data or the training data, by setting the appropriate flags. The method returns a tuple containing four lists: the (per-epoch) costs on the evaluation data, the accuracies on the evaluation data, the costs on the training data, and the accuracies on the training data. All values are evaluated at the end of each training epoch. So, for example, if we train for 30 epochs, then the first element of the tuple will be a 30-element list containing the cost on the evaluation data at the end of each epoch. Note that the lists are empty if the corresponding flag is not set.

**update\_mini\_batch** (*mini\_batch*, *eta*, *lmbda*, *n*)

Update the network's weights and biases by applying gradient descent using backpropagation to a single mini batch. The *mini\_batch* is a list of tuples (*x*, *y*), *eta* is the learning rate, *lmbda* is the regularization parameter, and *n* is the total size of the training data set.

**backprop** (*x*, *y*)

Return a tuple (*nabla\_b*, *nabla\_w*) representing the gradient for the cost function *C\_x*. *nabla\_b* and *nabla\_w* are layer-by-layer lists of numpy arrays, similar to *self.biases* and *self.weights*.

**accuracy** (*data*, *convert*=False)

Return the number of inputs in *data* for which the neural network outputs the correct result. The neural network's output is assumed to be the index of whichever neuron in the final layer has the highest activation.

The flag *convert* should be set to False if the data set is validation or test data (the usual case), and to True if the data set is the training data. The need for this flag arises due to differences in the way the results *y* are represented in the different data sets. In particular, it flags whether we need to convert between the different representations. It may seem strange to use different representations for the different data sets. Why not use the same representation for all three data sets? It's done for efficiency reasons – the program usually evaluates the cost on the training data and the accuracy on other data sets. These are different types of computations, and using different representations speeds things up. More details on the representations can be found in *mnist\_loader.load\_data\_wrapper*.

**total\_cost** (*data*, *lmbda*, *convert*=False)

Return the total cost for the data set *data*. The flag *convert* should be set to False if the data set is the training data (the usual case), and to True if the data set is the validation or test data. See comments on the similar (but reversed) convention for the *accuracy* method, above.

**save** (*filename*)

Save the neural network to the file *filename*.

**network2.load** (*filename*)

Load a neural network from the file *filename*. Returns an instance of *Network*.

**network2.vectorized\_result** (*j*)

Return a 10-dimensional unit vector with a 1.0 in the *j*'th position and zeroes elsewhere. This is used to convert a digit (0...9) into a corresponding desired output from the neural network.

**network2.sigmoid** (*z*)

The sigmoid function.

**network2.sigmoid\_prime** (*z*)

Derivative of the sigmoid function.

## 1.9 network3 module

### 1.10 test module

Testing code for different neural network configurations. Adapted for Python 3.5.2

**Usage in shell:** python3.5 test.py

**Network (network.py and network2.py) parameters:** 2nd param is epochs count 3rd param is batch size 4th param is learning rate (eta)

**Author:** Michał Dobrzański, 2016 [dobrzanski.michal.daniel@gmail.com](mailto:dobrzanski.michal.daniel@gmail.com)

```
test.main()
```

### 1.11 test1 module

```
test1.main()
```

```
test1.visualizeResults (dataSet, fname='backups/latest.pkl', saveSuccesses=False)
```

```
test1.train()
```

```
test1.test()
```

## 1.12 tests package

### 1.12.1 Submodules

#### 1.12.2 tests.conftest module

`tests.conftest.get_dummy_training_data (inputSize, outputSize, count)`  
creates a list of randomized (dummy) training data (X, Y) of the desired dimentions.

#### 1.12.3 tests.test\_network1 module

##### `test_network1.py`

unit test mynet.py against network.py (the provided implementation)

`tests.test_network1.test_feed_forward()`  
test that feedforward() behaves the same across the two implemenations

`tests.test_network1.test_backprop()`  
test that backprop() behaves the same across the two implemenations

`tests.test_network1.test_update_mini_batch()`  
test that update\_mini\_batch() behaves the same across the two implemenations

`tests.test_network1.test_SGD()`  
test that SGD() behaves the same across the two implemenations

#### 1.12.4 Module contents



## INDICES AND TABLES

- `genindex`
- `modindex`
- `search`





## PYTHON MODULE INDEX

### b

bootyNet, 1

### e

expand\_mnist, 1

### m

mnist\_average\_darkness, 1

mnist\_loader, 2

mnist\_svm, 3

mynet, 3

### n

network, 5

network2, 6

### t

test, 8

test1, 8

tests, 9

tests.conftest, 8

tests.test\_network1, 8



## Symbols

`__init__()` (*mynet.Network* method), 3  
`__init__()` (*network.Network* method), 5  
`__init__()` (*network2.Network* method), 6

## A

`accuracy()` (*network2.Network* method), 7  
`avg_darknesses()` (in *mnist\_average\_darkness*), 1

## B

`backprop()` (*mynet.Network* method), 4  
`backprop()` (*network.Network* method), 5  
`backprop()` (*network2.Network* method), 7  
`biases` (*mynet.Network* attribute), 3  
`bootyNet`  
*module*, 1

## C

`cost_derivative()` (*mynet.Network* method), 4  
`cost_derivative()` (*network.Network* method), 5  
`CrossEntropyCost` (class in *network2*), 6

## D

`default_weight_initializer()` (*network2.Network* method), 6  
`delta()` (*network2.CrossEntropyCost* static method), 6  
`delta()` (*network2.QuadraticCost* static method), 6

## E

`evaluate()` (*network.Network* method), 5  
`expand_mnist`  
*module*, 1

## F

`feedforward()` (*mynet.Network* method), 4  
`feedforward()` (*network.Network* method), 5  
`feedforward()` (*network2.Network* method), 6  
`fn()` (*network2.CrossEntropyCost* static method), 6  
`fn()` (*network2.QuadraticCost* static method), 6

## G

`get_dummy_training_data()` (in *tests.conftest*), 8  
`getConvergeRate()` (in *module bootyNet*), 1  
`getOutput()` (*mynet.Network* method), 4  
`goon()` (in *module bootyNet*), 1  
`guess_digit()` (in *module mnist\_average\_darkness*), 2

## L

`large_weight_initializer()` (*network2.Network* method), 6  
`load()` (in *module network2*), 7  
`load()` (*mynet.Network* method), 4  
`load()` (*network.Network* method), 5  
`load_data()` (in *module mnist\_loader*), 2  
`load_data_wrapper()` (in *module mnist\_loader*), 2

## M

`main()` (in *module bootyNet*), 1  
`main()` (in *module expand\_mnist*), 1  
`main()` (in *module mnist\_average\_darkness*), 1  
`main()` (in *module test*), 8  
`main()` (in *module test1*), 8  
`mnist_average_darkness`  
*module*, 1  
`mnist_loader`  
*module*, 2  
`mnist_svm`  
*module*, 3  
*module*  
*bootyNet*, 1  
*expand\_mnist*, 1  
*mnist\_average\_darkness*, 1  
*mnist\_loader*, 2  
*mnist\_svm*, 3  
*mynet*, 3  
*network*, 5  
*network2*, 6  
*test*, 8  
*test1*, 8  
*tests*, 9

tests.conftest, 8  
tests.test\_network1, 8  
mynet  
    module, 3

## N

network  
    module, 5  
Network (class in mynet), 3  
Network (class in network), 5  
Network (class in network2), 6  
network2  
    module, 6  
num\_layers (mynet.Network attribute), 3

## Q

QuadraticCost (class in network2), 6

## S

save() (mynet.Network method), 4  
save() (network.Network method), 5  
save() (network2.Network method), 7  
SGD() (mynet.Network method), 4  
SGD() (network.Network method), 5  
SGD() (network2.Network method), 6  
sigmoid() (in module network), 5  
sigmoid() (in module network2), 7  
sigmoid() (mynet.Network method), 4  
sigmoid\_prime() (in module network), 5  
sigmoid\_prime() (in module network2), 7  
sigmoid\_prime() (mynet.Network method), 4  
sizes (mynet.Network attribute), 3  
svm\_baseline() (in module mnist\_svm), 3

## T

test  
    module, 8  
test() (in module test1), 8  
test() (mynet.Network method), 4  
test1  
    module, 8  
test\_backprop() (in module tests.test\_network1), 8  
test\_feed\_forward() (in module tests.test\_network1), 8  
test\_SGD() (in module tests.test\_network1), 8  
test\_update\_mini\_batch() (in module tests.test\_network1), 8  
tests  
    module, 9  
tests.conftest  
    module, 8  
tests.test\_network1  
    module, 8  
total\_cost() (network2.Network method), 7

train() (in module test1), 8

## U

update\_mini\_batch() (network.Network method), 5  
update\_mini\_batch() (network2.Network method), 7  
updateMiniBatch() (mynet.Network method), 4

## V

vectorized\_result() (in module mnist\_loader), 2  
vectorized\_result() (in module network2), 7  
vectorToImage() (in module mnist\_loader), 2  
visualizeResults() (in module bootyNet), 1  
visualizeResults() (in module test1), 8

## W

weights (mynet.Network attribute), 3