Exercise 6: Solution

Activation Functions

Relu– Forward

```python
def forward(self, x):
    """
    :param x: Inputs, of any shape

    :return out: Output, of the same shape as x
    :return cache: Cache, for backward computation, of the same shape as x
    """
    outputs = None
    cache = None
    ########################################################################
    # TODO:                                                                #
    # Implement the forward pass of Relu activation function               #
    ########################################################################
    outputs = np.maximum(x, 0)
    cache = x
    ########################################################################
    #                            END OF YOUR CODE                          #
    ########################################################################
    return outputs, cache
```

Remark: when an element of input $x_{ij} > 0$, output x_{ij}, else output 0.

Relu– Backward

```
def backward(self, dout, cache):
    """
    :return: dx: the gradient w.r.t. input X, of the same shape as X
    """
    dx = None
    ########################################################################
    # TODO:                                                                #
    # Implement the backward pass of Relu activation function              #
    ########################################################################
    x = cache
    dx = dout
    # if x > 0, the gradient is 1, else 0.
    dx[x < 0] = 0
    ########################################################################
    #                           END OF YOUR CODE                           #
    ########################################################################
    return dx
```

Remark:
If the cache $x_{ij} \geq 0$, the gradient accordingly is 1, else 0.
Don't forget to multiply the upstreaming gradient.

LeakyRelu – Forward

```
def forward(self, x):
    """
    :param x: Inputs, of any shape

    :return out: Output, of the same shape as x
    :return cache: Cache, for backward computation, of the same shape as x
    """
    outputs = None
    cache = None
    ########################################################################
    # TODO:                                                                #
    # Implement the forward pass of LeakyRelu activation function          #
    ########################################################################
    cache = x
    outputs = x
    outputs[x <= 0] *= self.slope
    ########################################################################
    #                           END OF YOUR CODE                           #
    ########################################################################
    return outputs, cache
```

Remark:
What is different from Relu is, when input $x_{ij} \leq 0$, output is not 0, but $x_{ij} * slope$(0.01 by default).

LeakyRelu – Backward

```python
def backward(self, dout, cache):
    """
    :return: dx: the gradient w.r.t. input X, of the same shape as X
    """
    dx = None
    ########################################################################
    # TODO:                                                                #
    # Implement the backward pass of LeakyRelu activation function        #
    ########################################################################
    x = cache
    dx = dout
    dx[x <= 0] *= self.slope
    ########################################################################
    #                           END OF YOUR CODE                           #
    ########################################################################
    return dx
```

Remark:
What is different from Relu is, when the cache $x_{ij} \leq 0$, the gradient is not 0 but the slope.

Tanh – Forward

```python
def forward(self, x):
    """
    :param x: Inputs, of any shape

    :return out: Output, of the same shape as x
    :return cache: Cache, for backward computation, of the same shape as x
    """
    outputs = None
    cache = None
    ########################################################################
    # TODO:                                                                #
    # Implement the forward pass of Tanh activation function               #
    ########################################################################
    outputs = (np.exp(x) - np.exp(-x)) / (np.exp(x) + np.exp(-x))
    cache = outputs
    ########################################################################
    #                           END OF YOUR CODE                           #
    ########################################################################
    return outputs, cache
```

Remark:
Forward pass of Tanh is

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Optional:
You may also restore input x as cache.

Tanh – Backward

```python
def backward(self, dout, cache):
    """
    :return: dx: the gradient w.r.t. input X, of the same shape as X
    """
    dx = None
    ########################################################################
    # TODO:                                                                #
    # Implement the backward pass of Tanh activation function              #
    ########################################################################
    x = cache
    dx = 1 - x ** 2
    dx = dx * dout
    ########################################################################
    #                           END OF YOUR CODE                           #
    ########################################################################
    return dx
```

Remark:
The backward pass of Tanh is

$$\frac{dy}{dx} = \frac{4}{(e^x + e^{-x})^2}$$
$$= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2$$

Test the activation functions!

```
In [1]: %load_ext autoreload
        %autoreload 2
```

```
In [2]: from exercise_code.tests.layer_tests import *

        print(ReluTest()())
        print()
        print(LeakyReluTest()())
        print()
        print(TanhTest()())
```

```
ReluForwardTest passed.
ReluBackwardTest passed.
Congratulations you have passed all the unit tests!!! Tests passed: 2/2
(0, 2)

LeakyReluForwardTest passed.
LeakyReluBackwardTest passed.
Congratulations you have passed all the unit tests!!! Tests passed: 2/2
(0, 2)

TanhForwardTest passed.
TanhBackwardTest passed.
Congratulations you have passed all the unit tests!!! Tests passed: 2/2
(0, 2)
```

Random Search

A feasible set of range of hyperparameters

In [14]:

```
from exercise_code.networks import MyOwnNetwork

best_model = ClassificationNet()
#best_model = MyOwnNetwork()

########################################################################
# TODO:                                                                #
# Implement your own neural network and find suitable hyperparameters  #
# Be sure to edit the MyOwnNetwork class in the following code snippet #
# to upload the correct model!                                         #
########################################################################
from exercise_code.hyperparameter_tuning import random_search

best_model, results = random_search(dataloaders['train'], dataloaders['val'],
                                    random_search_spaces = {
                                      "learning_rate": ([1e-3, 1e-4], 'log'),
                                      "lr_decay": ([0.8, 0.9], 'float'),
                                      "reg": ([1e-4, 1e-6], "log"),
                                      "std": ([1e-4, 1e-6], "log"),
                                      "hidden_size": ([50, 100], "int"),
                                      "num_layer": ([2], "int"),
                                      "activation": ([Relu()], "item"),
                                      "optimizer": ([Adam], "item"),
                                      "loss_func": ([CrossEntropyFromLogits()], "item")
                                     }, num_search = 5, epochs=20, patience=5,
                                    model_class=ClassificationNet)
########################################################################
#                           END OF YOUR CODE                           #
########################################################################
```

Pick the best set of hyperparameters

```
Search done. Best Val Loss = 1.4614823323760282
Best Config: {'learning_rate': 0.0009363255745516442, 'lr_decay': 0.8106866888065208, 'reg': 3.5115962843695404e-05, 'std': 1.0074810757234067e-06, 'hidden_size': 96, 'num_layer': 2, 'activation': <exercise_code.networks.layer.Relu object at 0x7f3a256d52b0>, 'optimizer': <class 'exercise_code.networks.optimizer.Adam'>, 'loss_func': <exercise_code.networks.loss.CrossEntropyFromLogits object at 0x7f3a4cb2da00>}
```

Checking the validation accuracy

In [15]:
```
labels, pred, acc = best_model.get_dataset_prediction(dataloaders['train'])
print("Train Accuracy: {}%".format(acc*100))
labels, pred, acc = best_model.get_dataset_prediction(dataloaders['val'])
print("Validation Accuracy: {}%".format(acc*100))
```

```
Train Accuracy: 57.85590277777778%
Validation Accuracy: 49.23878205128205%
```

In [16]:
```
# comment this part out to see your model's performance on the test set.

labels, pred, acc = best_model.get_dataset_prediction(dataloaders['test'])
print("Test Accuracy: {}%".format(acc*100))
```

```
Test Accuracy: 49.318910256410255%
```

Questions? Moodle

☺