

Neural Networks

Ghazal Kalhor

Abstract — In this computer assignment, we want to classify given photos from lungs of humans into three classes using Neural Networks that we learnt in Artificial Intelligence. We will build our neural network using Keras library. **Keywords** — Artificial Intelligence, Keras, Neural Networks

Introduction

The aim of this computer assignment is to make an optimal neural network in order to classify the photos of the humans' lung into Covid19 , Normal , and Penumonia classes.

Importing Libraries

In this part, some of the necessary libraries were imported in order to use their helpful functions.

```
In []:
import numpy as np
from tensorflow.keras import *
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report
```

Defining Constants

In this part, constant values are defined in order to make the code more readable and more flexible to change.

```
In []:
N_CLASSES = 3
TRAIN_DIR = "/content/Data/train"
TEST_DIR = "/content/Data/test"
```

Importing Data

In this part, file *xray.zip* is copied to the project directory, then we unzipped it and put its data into *train* and *test* sub-directories.

```
In []:
!cp "/content/drive/MyDrive/xray.zip" .
```

```
In []:
!unzip xray.zip
```

Archive: xray.zip
 inflating: Data/test/COVID19/COVID19(460).jpg
 inflating: Data/test/COVID19/COVID19(461).jpg
 inflating: Data/test/COVID19/COVID19(462).jpg
 inflating: Data/test/COVID19/COVID19(463).jpg
 inflating: Data/test/COVID19/COVID19(464).jpg
 inflating: Data/test/COVID19/COVID19(465).jpg
 inflating: Data/test/COVID19/COVID19(466).jpg
 inflating: Data/test/COVID19/COVID19(467).jpg
 inflating: Data/test/COVID19/COVID19(468).jpg
 inflating: Data/test/COVID19/COVID19(469).jpg
 inflating: Data/test/COVID19/COVID19(470).jpg
 inflating: Data/test/COVID19/COVID19(471).jpg
 inflating: Data/test/COVID19/COVID19(472).jpg
 inflating: Data/test/COVID19/COVID19(473).jpg
 inflating: Data/test/COVID19/COVID19(474).jpg
 inflating: Data/test/COVID19/COVID19(475).jpg
 inflating: Data/test/COVID19/COVID19(476).jpg
 inflating: Data/test/COVID19/COVID19(477).jpg
 inflating: Data/test/COVID19/COVID19(478).jpg
 inflating: Data/test/COVID19/COVID19(479).jpg

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

In this part, we read test and train data, using `image.ImageDataGenerator()` method from keras library. Then we performed some preprocessing on the images. Moreover, we set the batch-size to 32 for the training process.

In []:

```
data_generator = preprocessing.image.ImageDataGenerator()
```

In []:

```
def get_data(data_generator, file_path):  
    data = data_generator.flow_from_directory(  
        file_path,  
        target_size = (80, 80),  
        color_mode = "grayscale",  
        batch_size = 32  
    )  
    return data
```

In this part, we printed the number of instances of each class in test and train data.

In []:

```
train_data = get_data(data_generator, TRAIN_DIR)
```

Found 5144 images belonging to 3 classes.

In []:

```
test_data = get_data(data_generator, TEST_DIR)
```

Found 1288 images belonging to 3 classes.

Question 2: Visualization

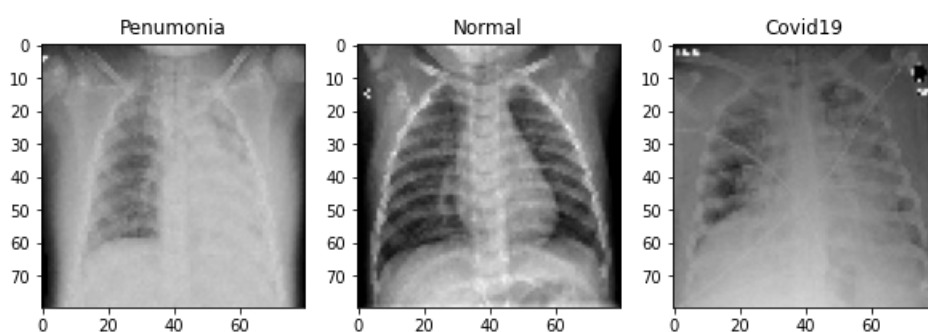
Plotting a Sample of each Class

In []:

```
images, labels = train_data.next()
```

In []:

```
fig, ax = plot_container = plt.subplots(1, 3)  
fig.set_size_inches(10, 4)  
  
label_dict = {0:'Covid19', 1:'Normal', 2:'Penumonia'}  
  
cursor = 0  
seen = list()  
  
for image, label in zip(images, labels):  
    class_label = backend.argmax(label).numpy()  
    if class_label not in seen:  
        ax[cursor].imshow(image.reshape(80, 80), cmap='gray')  
        ax[cursor].set_title(label_dict[class_label])  
        seen.append(class_label)  
        cursor += 1
```



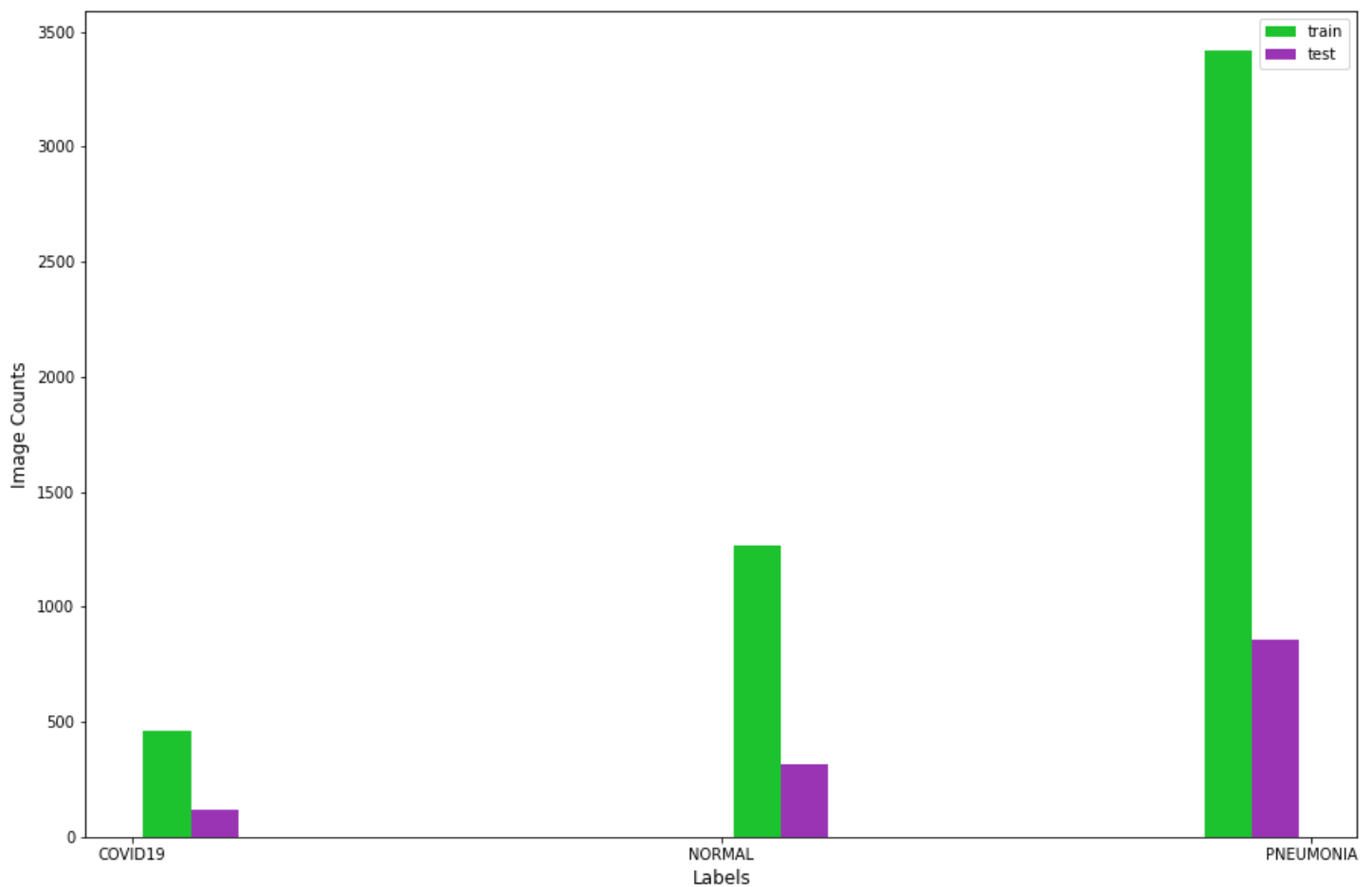
Plotting Frequency of each Class

In []:

```
fig = plt.figure(figsize=(15, 10))
labels = (train_data.classes, test_data.classes)
plt.hist(labels, label=['train', 'test'], color=['#1cc32f', '#9a34b5'])
plt.xlabel('Labels', fontsize = 12)
plt.ylabel('Image Counts', fontsize = 12)
plt.suptitle('Data Frequency of each Class', fontsize = 15)
plt.legend()
loc = list(train_data.class_indices.values())
my_xticks = list(train_data.class_indices.keys())
plt.xticks(loc, my_xticks)
plt.show()
```

/usr/local/lib/python3.6/dist-packages/numpy/core/_asarray.py:83: VisibleDeprecationWarning: Creating an ndarray from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths or shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarray
return array(a, dtype, copy=False, order=order)

Data Frequency of each Class



Question 3

In this part, we implemented necessary functions to make a neural network with and without dropout layers. Also, we set default values for some hyperparameters based on the project description.

In []:

```
def make_feedforward_network(neurons_1st_dense, neurons_2nd_dense, neurons_3rd_dense, activation='relu',
                             optimizer=optimizers.SGD(), loss=losses.categorical_crossentropy,
                             regularizer=None):
    inp = layers.Input(shape=(80, 80, 1))
    out = layers.Flatten()(inp)
    out = layers.Dense(neurons_1st_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dense(neurons_2nd_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dense(neurons_3rd_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dense(N_CLASSES, activation="softmax")(out)
```

```
model = models.Model(inputs=inp, outputs=out)
model.compile(optimizer=optimizer, loss=loss, metrics=['accuracy'])
return model
```

In []:

```
def make_feedforward_network_with_dropout(neurons_1st_dense, neurons_2nd_dense, neurons_3rd_dense, dropout, \
    activation='relu', optimizer=optimizers.SGD(), \
    loss=losses.categorical_crossentropy, regularizer=None):
    inp = layers.Input(shape=(80, 80, 1))
    out = layers.Flatten()(inp)
    out = layers.Dense(neurons_1st_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dropout(dropout)(out)
    out = layers.Dense(neurons_2nd_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dropout(dropout)(out)
    out = layers.Dense(neurons_3rd_dense, activation=activation, kernel_regularizer=regularizer)(out)
    out = layers.Dropout(dropout)(out)
    out = layers.Dense(N_CLASSES, activation="softmax")(out)

    model = models.Model(inputs=inp, outputs=out)
    model.compile(optimizer=optimizer, loss=loss, metrics=['accuracy'])
    return model
```

Constructing FeedForward Neural Network

In []:

```
relu_model = make_feedforward_network(4800, 3200, 1600)
```

Parameters of the Model

In this part, we use `summary()` method to check the parameters of our model.

Each row represents a layer with each named uniquely such that we can refer to these layers without any ambiguity. As you can see, the layers that we added to the model in the previous code snippet are all included in the output.

Each layer has an output and its shape is shown in the **Output Shape** column Each layer's output becomes the input for the subsequent layer.

The **Param #** column shows the number of parameters that are trained for each layer.

The total number of parameters is shown at the end, which is equal to the number of trainable and non-trainable parameters. In this model, all the parameters are trainable.

In []:

```
relu_model.summary()
```

Model: "model"

| Layer (type) | Output Shape | Param # |
|------------------------------|---------------------|----------|
| ===== | | |
| input_1 (InputLayer) | [(None, 80, 80, 1)] | 0 |
| flatten (Flatten) | (None, 6400) | 0 |
| dense (Dense) | (None, 4800) | 30724800 |
| dense_1 (Dense) | (None, 3200) | 15363200 |
| dense_2 (Dense) | (None, 1600) | 5121600 |
| dense_3 (Dense) | (None, 3) | 4803 |
| ===== | | |
| Total params: 51,214,403 | | |
| Trainable params: 51,214,403 | | |
| Non-trainable params: 0 | | |

The number of parameters for each layer can be computed using the following formula:

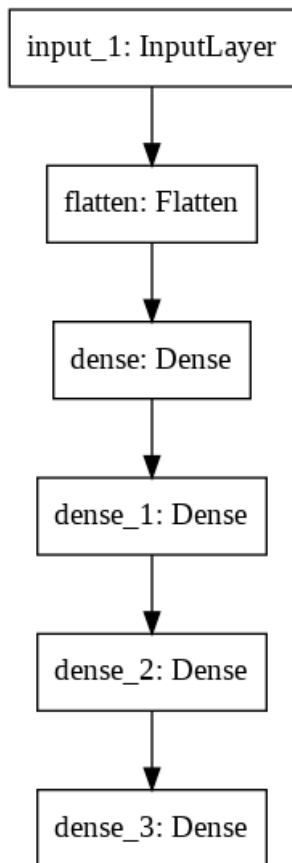
$$\# \text{ of parameters} = (\# \text{ of inputs} + 1) (\# \text{ of outputs})^*$$

1 in the above formula refers to the bias term and the product shows the number of edges from the previous layer to the current layer. We need to compute the weights of these edges in our training process.

In []:

```
utils.plot_model(relu_model)
```

Out[]:



Question 4

Defining Methods for Training

In []:

```
def get_labels(data):
    target_labels = []
    for i in range(int(np.ceil(len(data.classes)/data.batch_size))):
        images, labels = data.next()
        for label in labels:
            target_labels.append(backend.argmax(label))
        target_labels = np.array(target_labels)
    return target_labels
```

In []:

```
def print_metrics(true_outputs, expected_outputs):
    true_outputs = np.array(backend.argmax(true_outputs))
    print(classification_report(expected_outputs, true_outputs))
```

In []:

```
def predict(model, test_data):
    expected_outputs = get_labels(test_data)
    true_outputs = model.predict(test_data)
    print_metrics(true_outputs, expected_outputs)
```

In []:

```
def plot_loss(history):
    fig = plt.figure(figsize=(15, 10))
    plt.plot(history.history['loss'], linewidth = 3, color = '#e500a9')
    plt.plot(history.history['val_loss'], linewidth = 3, color = '#9a34b5')
    plt.title('Model Loss', fontsize = 15)
    plt.ylabel('Loss', fontsize = 12)
```

```
plt.xlabel('Epoch', fontsize = 12)
plt.legend(['Train', 'Test'], loc='upper left')
plt.show()
```

Training with ReLU Activation Function

When we use this activation function, we face with the nan loss values and this happend because we did not normalize the photos in the previous parts. Moreover, we have died relu and the accuracy of our model is not updating.

In []:

```
history = relu_model.fit(train_data, validation_data=test_data, epochs=10)
```

```
Epoch 1/10
161/161 [=====] - 170s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 2/10
161/161 [=====] - 161s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 3/10
161/161 [=====] - 165s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 4/10
161/161 [=====] - 164s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 5/10
161/161 [=====] - 162s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 6/10
161/161 [=====] - 161s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 7/10
161/161 [=====] - 163s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 8/10
161/161 [=====] - 162s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 9/10
161/161 [=====] - 164s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
Epoch 10/10
161/161 [=====] - 168s 1s/step - loss: nan - accuracy: 0.0894 - val_loss: nan - val_accuracy: 0.0901
```

In []:

```
predict(relu_model, test_data)
```

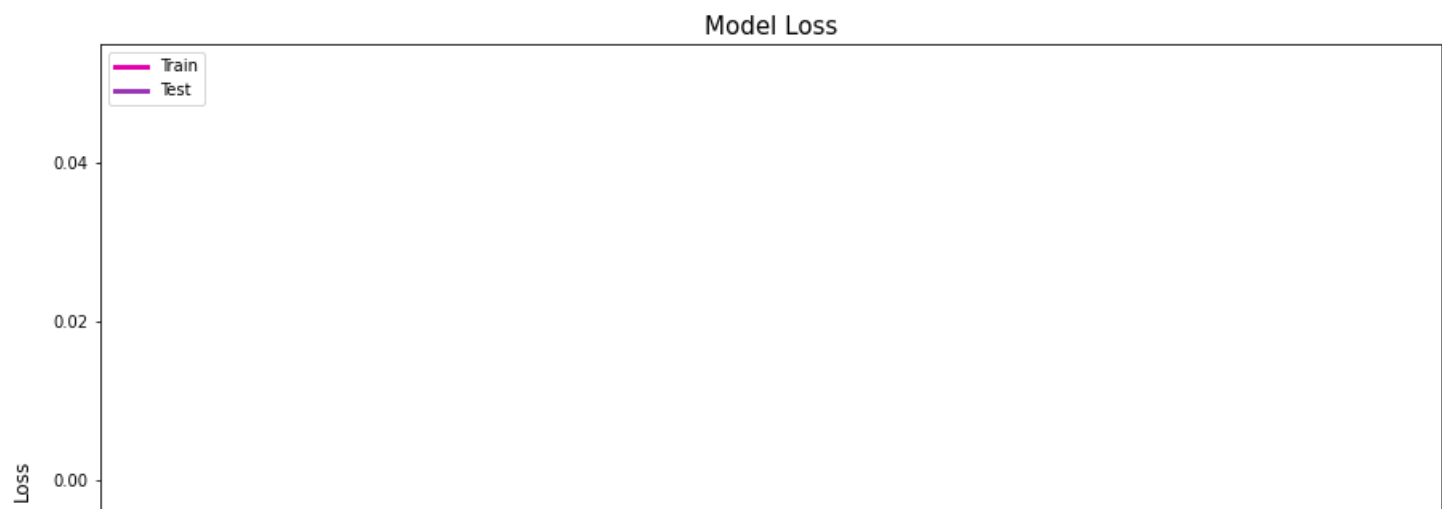
| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.09 | 1.00 | 0.17 | 116 |
| 1 | 0.00 | 0.00 | 0.00 | 317 |
| 2 | 0.00 | 0.00 | 0.00 | 855 |
| accuracy | | | 0.09 | 1288 |
| macro avg | 0.03 | 0.33 | 0.06 | 1288 |
| weighted avg | 0.01 | 0.09 | 0.01 | 1288 |

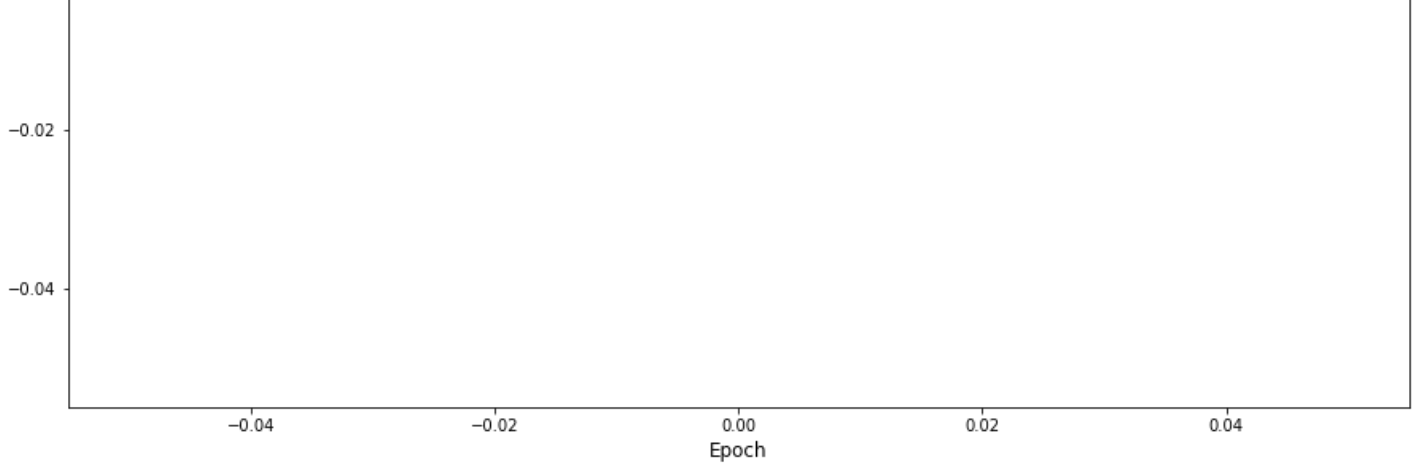
/usr/local/lib/python3.6/dist-packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```
_warn_prf(average, modifier, msg_start, len(result))
```

In []:

```
plot_loss(history)
```





Training with Tanh Activation Function

When we use the tanh activation function, gradients get smaller and smaller during the time. Therefore, neurons of the first layers do the learning process slower than neurons of the last layers. This can make the learning process longer. As a result, the accuracy of the model will be reduced.

In []:

```
tanh_model = make_feedforward_network(4800, 3200, 1600, activation='tanh')
```

In []:

```
history = tanh_model.fit(train_data, validation_data=test_data, epochs=10)
```

Epoch 1/10
161/161 [=====] - 171s 1s/step - loss: 2.3961 - accuracy: 0.5267 - val_loss: 0.9371 - val_accuracy: 0.6661
Epoch 2/10
161/161 [=====] - 165s 1s/step - loss: 1.2054 - accuracy: 0.5810 - val_loss: 1.0981 - val_accuracy: 0.6638
Epoch 3/10
161/161 [=====] - 163s 1s/step - loss: 0.9745 - accuracy: 0.6278 - val_loss: 1.0855 - val_accuracy: 0.6638
Epoch 4/10
161/161 [=====] - 164s 1s/step - loss: 0.9137 - accuracy: 0.6491 - val_loss: 0.9932 - val_accuracy: 0.6638
Epoch 5/10
161/161 [=====] - 163s 1s/step - loss: 0.9174 - accuracy: 0.6262 - val_loss: 1.0112 - val_accuracy: 0.6638
Epoch 6/10
161/161 [=====] - 163s 1s/step - loss: 0.9061 - accuracy: 0.6306 - val_loss: 0.9264 - val_accuracy: 0.6638
Epoch 7/10
161/161 [=====] - 162s 1s/step - loss: 0.8706 - accuracy: 0.6356 - val_loss: 0.8564 - val_accuracy: 0.6638
Epoch 8/10
161/161 [=====] - 162s 1s/step - loss: 0.9044 - accuracy: 0.6366 - val_loss: 0.8893 - val_accuracy: 0.6638
Epoch 9/10
161/161 [=====] - 162s 1s/step - loss: 0.8859 - accuracy: 0.6449 - val_loss: 0.8556 - val_accuracy: 0.6638
Epoch 10/10
161/161 [=====] - 162s 1s/step - loss: 0.8929 - accuracy: 0.6414 - val_loss: 0.9008 - val_accuracy: 0.6638

In []:

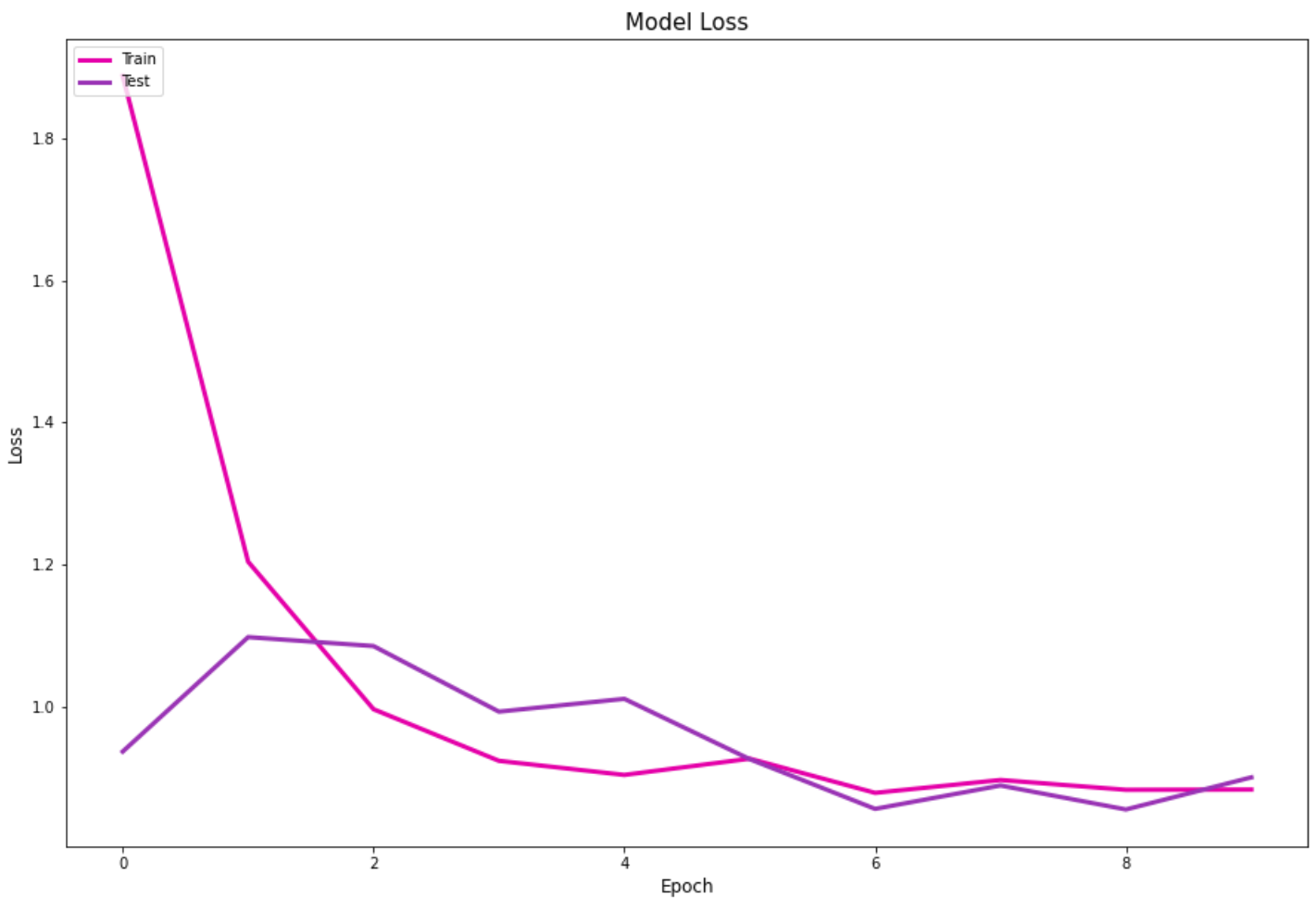
```
predict(tanh_model, test_data)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.00 | 0.00 | 0.00 | 116 |
| 1 | 0.00 | 0.00 | 0.00 | 317 |
| 2 | 0.66 | 1.00 | 0.80 | 855 |
| accuracy | | | 0.66 | 1288 |
| macro avg | 0.22 | 0.33 | 0.27 | 1288 |
| weighted avg | 0.44 | 0.66 | 0.53 | 1288 |

/usr/local/lib/python3.6/dist-packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))

In []:

```
plot_loss(history)
```



Comparison between Two Activation Functions

In both, our model is not updating and it seems that the process of training has been stopped. But the loss for the relu gets nan value and this refers to the definition of this function.

One technique to deal with this problem is to normalize the data before learning.

Question 5

In this part, we normalize data and set the number of hidden layers and the number of neurons in each layer to reach an optimal model. The F1-Score of our model is 0.96.

In []:

```
rescaled_data_generator = preprocessing.image.ImageDataGenerator(
    rescale=1. / 255
)
```

In []:

```
rescaled_train = get_data(rescaled_data_generator, TRAIN_DIR)
```

Found 5144 images belonging to 3 classes.

In []:

```
rescaled_test = get_data(rescaled_data_generator, TEST_DIR)
```

Found 1288 images belonging to 3 classes.

In []:

```
optimal_model = make_feedforward_network(4800, 3200, 1600)
history = optimal_model.fit(rescaled_train, validation_data=rescaled_test, epochs=10)
```

Epoch 1/10

4814/4814 samples, 1128 images, 1128 classes, 0.0000 loss, 0.7000 val loss, 0.0000 acc, 0.7400 val acc

161/161 [=====] - 166s 1s/step - loss: 0.6909 - accuracy: 0.7399 - val_loss: 0.6836 - val_accuracy: 0.7469
Epoch 2/10
161/161 [=====] - 167s 1s/step - loss: 0.3880 - accuracy: 0.8502 - val_loss: 0.2452 - val_accuracy: 0.9177
Epoch 3/10
161/161 [=====] - 168s 1s/step - loss: 0.3286 - accuracy: 0.8711 - val_loss: 0.2562 - val_accuracy: 0.8890
Epoch 4/10
161/161 [=====] - 168s 1s/step - loss: 0.2670 - accuracy: 0.8932 - val_loss: 0.2157 - val_accuracy: 0.9161
Epoch 5/10
161/161 [=====] - 169s 1s/step - loss: 0.2794 - accuracy: 0.8882 - val_loss: 0.1755 - val_accuracy: 0.9425
Epoch 6/10
161/161 [=====] - 169s 1s/step - loss: 0.2397 - accuracy: 0.9095 - val_loss: 0.4377 - val_accuracy: 0.8129
Epoch 7/10
161/161 [=====] - 168s 1s/step - loss: 0.2528 - accuracy: 0.9032 - val_loss: 0.1908 - val_accuracy: 0.9239
Epoch 8/10
161/161 [=====] - 168s 1s/step - loss: 0.2300 - accuracy: 0.9091 - val_loss: 0.2230 - val_accuracy: 0.9099
Epoch 9/10
161/161 [=====] - 167s 1s/step - loss: 0.2230 - accuracy: 0.9148 - val_loss: 0.3225 - val_accuracy: 0.8734
Epoch 10/10
161/161 [=====] - 167s 1s/step - loss: 0.2190 - accuracy: 0.9146 - val_loss: 0.1549 - val_accuracy: 0.9464

In []:

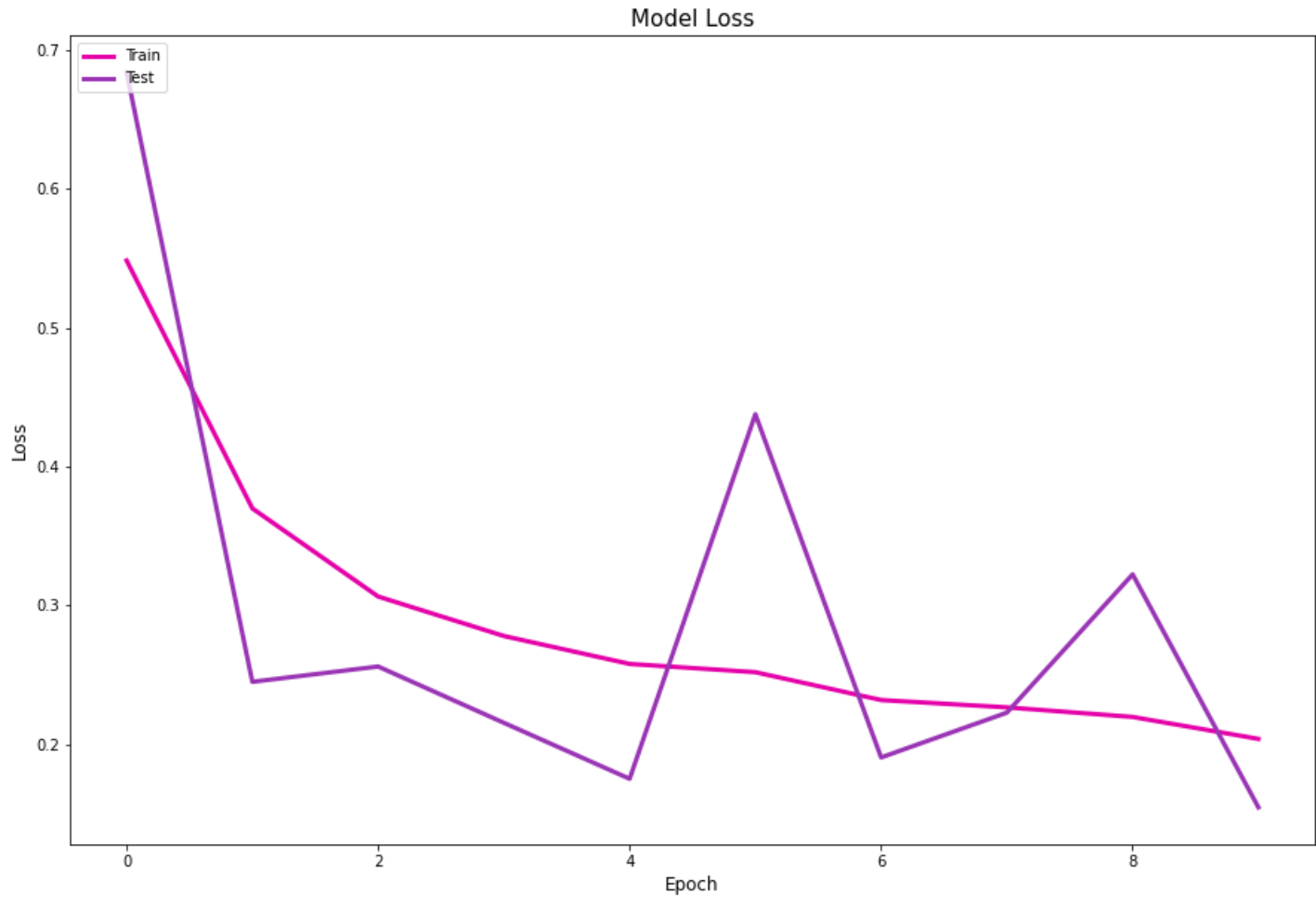
```
predict(optimal_model, rescaled_test)
```

| | precision | recall | f1-score | support |
|---|-----------|--------|----------|---------|
| 0 | 0.98 | 0.87 | 0.92 | 116 |
| 1 | 0.91 | 0.91 | 0.91 | 317 |
| 2 | 0.96 | 0.97 | 0.96 | 855 |

| | | | | |
|--------------|-----------|------|------|------|
| accuracy | 0.95 1288 | | | |
| macro avg | 0.95 | 0.92 | 0.93 | 1288 |
| weighted avg | 0.95 | 0.95 | 0.95 | 1288 |

In []:

```
plot_loss(history)
```



What is momentum in neural networks?

Momentum or SGD with momentum is a method which helps us to accelerate gradient vectors in the right directions. Momentum is an average of gradients that changes during the time and we use it to update weight in each step. Therefore, it leads to faster convergence. It is mostly used in neural networks considering the size of data in NNs makes a considerable time difference while training gradients. It can be used to handle noisy gradients. Moreover, it can handle extremely small gradients.

Is it always good to set momentum to a larger number?

Momentum or SGD with momentum is a method which helps us to accelerate gradient In fact, momentum makes each step of the learning process dependent on its previous. If we set momentum and learning rate to large values, we will have bigger steps, and based on the dependency that we mentioned above, the next steps will be larger and larger. As a result, weights won't be updated in a correct way, and they will never reach optimal values. Therefore, large momentum does not necessarily lead to higher accuracies.

Setting Momentum to 0.5

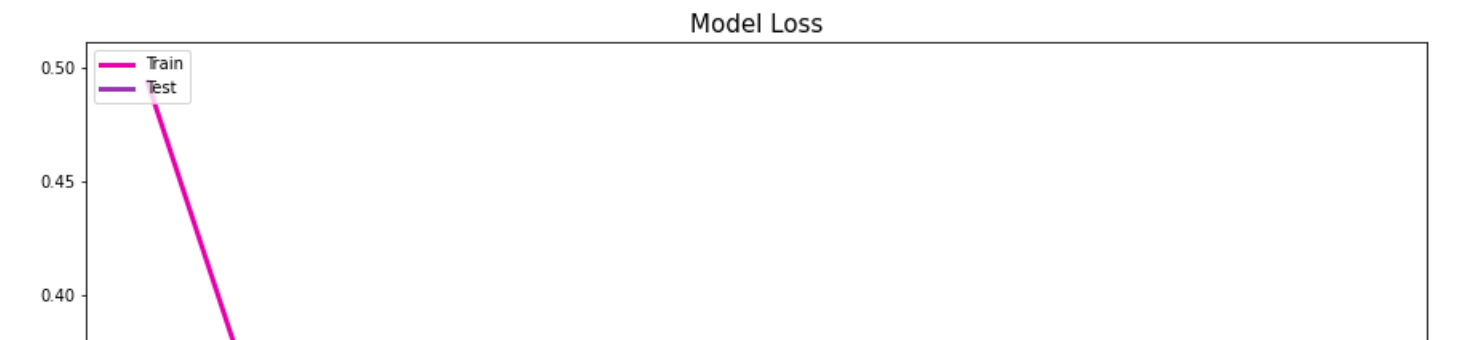
```
In []:
model_6_1 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.SGD(momentum=0.5))
history_6_1 = model_6_1.fit(rescaled_train, validation_data=rescaled_test, epochs=10)
```

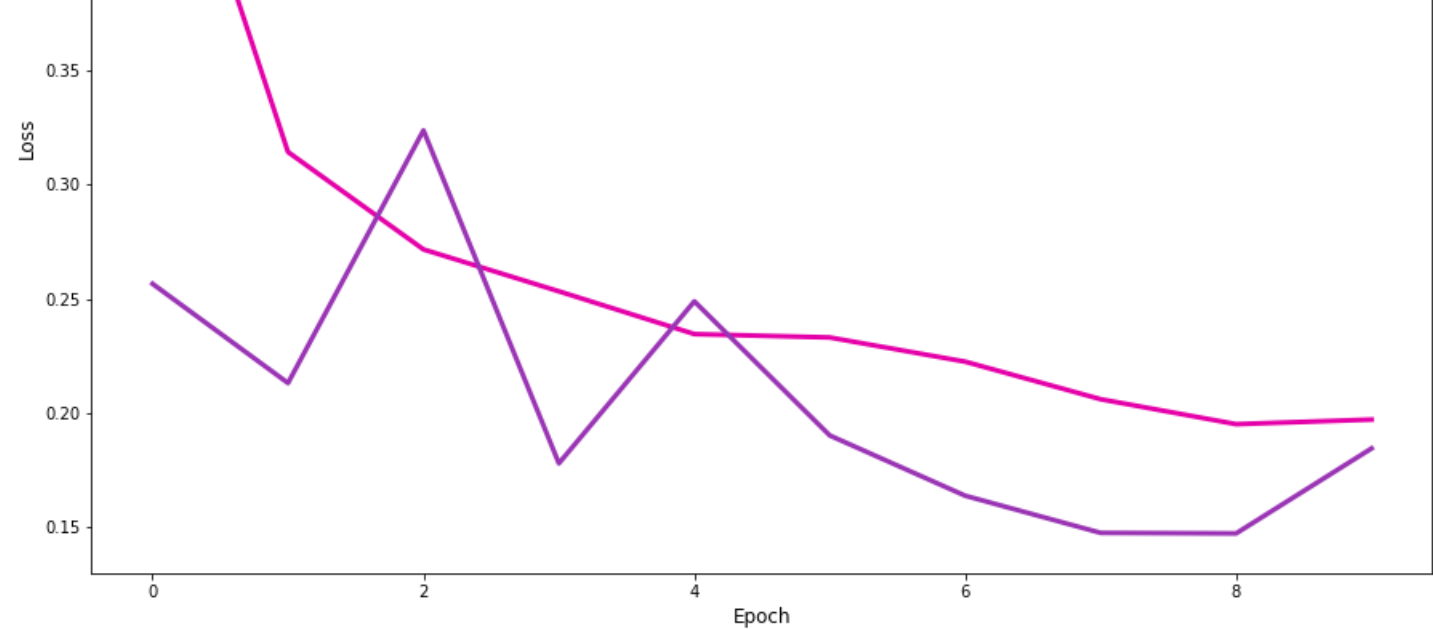
Epoch 1/10
161/161 [=====] - 146s 899ms/step - loss: 0.6972 - accuracy: 0.7242 - val_loss: 0.2565 - val_accuracy: 0.9006
Epoch 2/10
161/161 [=====] - 144s 897ms/step - loss: 0.3270 - accuracy: 0.8699 - val_loss: 0.2131 - val_accuracy: 0.9193
Epoch 3/10
161/161 [=====] - 145s 904ms/step - loss: 0.2706 - accuracy: 0.8901 - val_loss: 0.3237 - val_accuracy: 0.8672
Epoch 4/10
161/161 [=====] - 146s 909ms/step - loss: 0.2656 - accuracy: 0.9004 - val_loss: 0.1780 - val_accuracy: 0.9356
Epoch 5/10
161/161 [=====] - 145s 897ms/step - loss: 0.2284 - accuracy: 0.9139 - val_loss: 0.2489 - val_accuracy: 0.9053
Epoch 6/10
161/161 [=====] - 145s 900ms/step - loss: 0.2407 - accuracy: 0.9078 - val_loss: 0.1902 - val_accuracy: 0.9348
Epoch 7/10
161/161 [=====] - 146s 906ms/step - loss: 0.2466 - accuracy: 0.9084 - val_loss: 0.1639 - val_accuracy: 0.9425
Epoch 8/10
161/161 [=====] - 145s 900ms/step - loss: 0.1959 - accuracy: 0.9247 - val_loss: 0.1476 - val_accuracy: 0.9472
Epoch 9/10
161/161 [=====] - 145s 900ms/step - loss: 0.2007 - accuracy: 0.9207 - val_loss: 0.1474 - val_accuracy: 0.9488
Epoch 10/10
161/161 [=====] - 145s 903ms/step - loss: 0.2047 - accuracy: 0.9262 - val_loss: 0.1846 - val_accuracy: 0.9363

```
In []:
predict(model_6_1, rescaled_test)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.83 | 0.98 | 0.90 | 116 |
| 1 | 0.95 | 0.85 | 0.90 | 317 |
| 2 | 0.95 | 0.96 | 0.95 | 855 |
| accuracy | | | 0.94 | 1288 |
| macro avg | 0.91 | 0.93 | 0.92 | 1288 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1288 |

```
In []:
plot_loss(history_6_1)
```





Based on the result, this model is more optimal in comparison to the other model (with another values for momentum) and the loss decreases with higher rate.

Setting Momentum to 0.9

In []:

```
model_6_2 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.SGD(momentum=0.9))
history_6_2 = model_6_2.fit(rescaled_train, validation_data=rescaled_test, epochs=10)
```

Epoch 1/10
161/161 [=====] - 147s 914ms/step - loss: 0.6650 - accuracy: 0.7485 - val_loss: 0.2160 - val_accuracy: 0.9154
Epoch 2/10
161/161 [=====] - 144s 895ms/step - loss: 0.3101 - accuracy: 0.8804 - val_loss: 0.2032 - val_accuracy: 0.9208
Epoch 3/10
161/161 [=====] - 144s 892ms/step - loss: 0.2893 - accuracy: 0.8852 - val_loss: 0.2767 - val_accuracy: 0.8952
Epoch 4/10
161/161 [=====] - 144s 893ms/step - loss: 0.2778 - accuracy: 0.8925 - val_loss: 0.2195 - val_accuracy: 0.9177
Epoch 5/10
161/161 [=====] - 144s 891ms/step - loss: 0.2361 - accuracy: 0.9075 - val_loss: 0.1797 - val_accuracy: 0.9363
Epoch 6/10
161/161 [=====] - 143s 890ms/step - loss: 0.2572 - accuracy: 0.9009 - val_loss: 0.1876 - val_accuracy: 0.9262
Epoch 7/10
161/161 [=====] - 143s 891ms/step - loss: 0.2369 - accuracy: 0.9112 - val_loss: 0.3747 - val_accuracy: 0.8641
Epoch 8/10
161/161 [=====] - 143s 889ms/step - loss: 0.2220 - accuracy: 0.9154 - val_loss: 0.2936 - val_accuracy: 0.8680
Epoch 9/10
161/161 [=====] - 143s 888ms/step - loss: 0.2237 - accuracy: 0.9156 - val_loss: 0.2219 - val_accuracy: 0.9107
Epoch 10/10
161/161 [=====] - 143s 889ms/step - loss: 0.2037 - accuracy: 0.9183 - val_loss: 0.1695 - val_accuracy: 0.9410

In []:

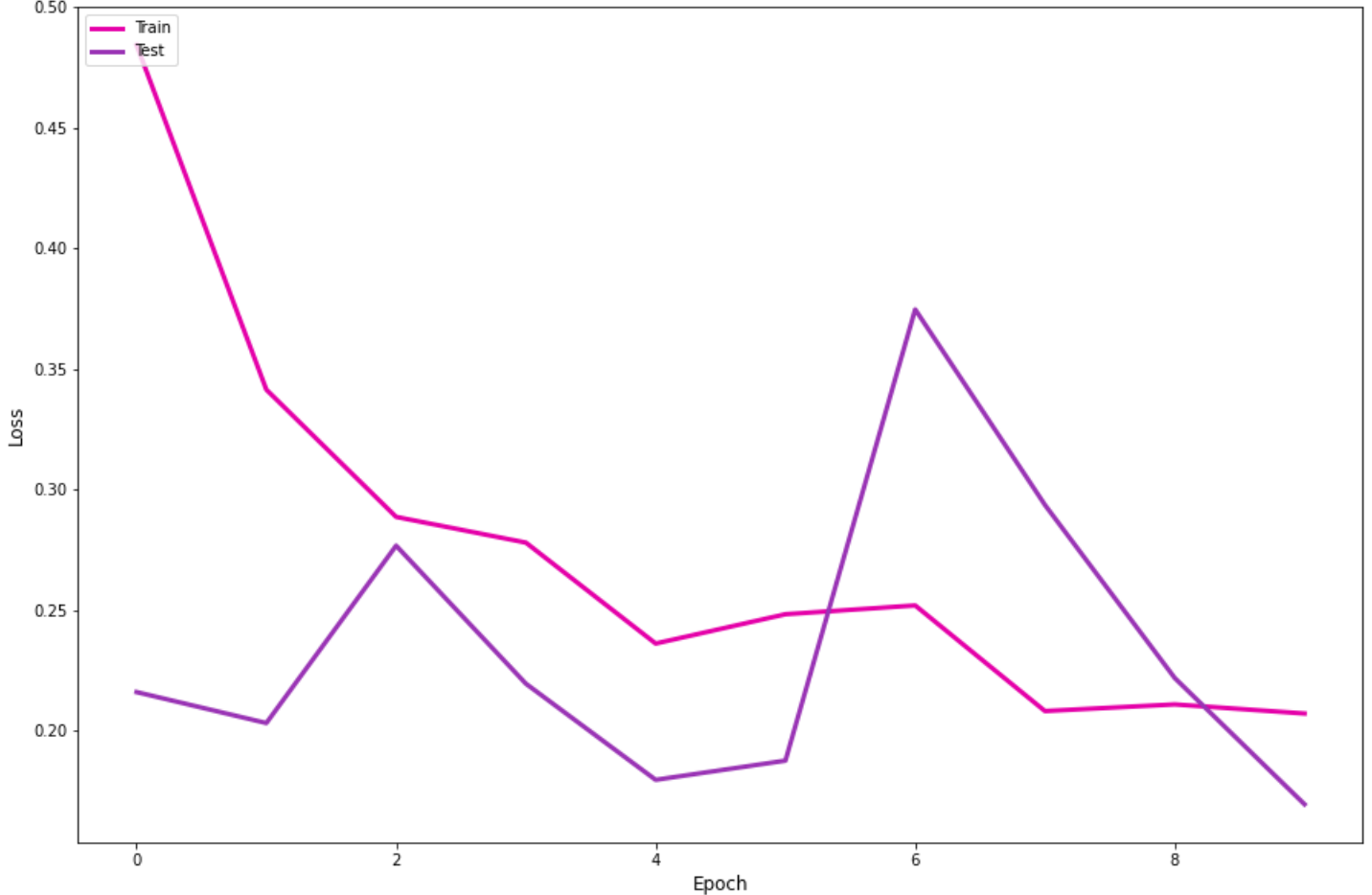
```
predict(model_6_2, rescaled_test)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.94 | 0.91 | 0.92 | 116 |
| 1 | 0.87 | 0.94 | 0.90 | 317 |
| 2 | 0.97 | 0.95 | 0.96 | 855 |
| accuracy | | | 0.94 | 1288 |
| macro avg | 0.93 | 0.93 | 0.93 | 1288 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1288 |

In []:

```
plot_loss(history_6_2)
```

Model Loss



Based on the result, using 0.9 for momentum does not have any positive effect and the rate of the changing in loss has been decreased.

Setting Momentum to 0.99

In []:

```
model_6_3 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.SGD(momentum=0.99))
history_6_3 = model_6_3.fit(rescaled_train, validation_data=rescaled_test, epochs=10)
```

```
Epoch 1/10
161/161 [=====] - 148s 916ms/step - loss: 0.8540 - accuracy: 0.6853 - val_loss: 0.2317 - val_accuracy: 0.9224
Epoch 2/10
161/161 [=====] - 147s 916ms/step - loss: 0.3961 - accuracy: 0.8543 - val_loss: 0.3337 - val_accuracy: 0.8719
Epoch 3/10
161/161 [=====] - 145s 901ms/step - loss: 0.4585 - accuracy: 0.8399 - val_loss: 0.3148 - val_accuracy: 0.8835
Epoch 4/10
161/161 [=====] - 145s 901ms/step - loss: 0.5077 - accuracy: 0.8301 - val_loss: 0.2747 - val_accuracy: 0.9371
Epoch 5/10
161/161 [=====] - 145s 901ms/step - loss: 0.4583 - accuracy: 0.8733 - val_loss: 0.4889 - val_accuracy: 0.8602
Epoch 6/10
161/161 [=====] - 143s 891ms/step - loss: 0.4640 - accuracy: 0.8477 - val_loss: 0.2731 - val_accuracy: 0.9309
Epoch 7/10
161/161 [=====] - 144s 899ms/step - loss: 0.3449 - accuracy: 0.9024 - val_loss: 0.5148 - val_accuracy: 0.8703
Epoch 8/10
161/161 [=====] - 144s 899ms/step - loss: 0.5021 - accuracy: 0.8432 - val_loss: 0.4781 - val_accuracy: 0.8540
Epoch 9/10
161/161 [=====] - 144s 895ms/step - loss: 0.4804 - accuracy: 0.8433 - val_loss: 0.8072 - val_accuracy: 0.7741
Epoch 10/10
161/161 [=====] - 144s 899ms/step - loss: 0.7433 - accuracy: 0.7686 - val_loss: 0.7684 - val_accuracy: 0.7943
```

In []:

```
predict(model_6_3, rescaled_test)
```

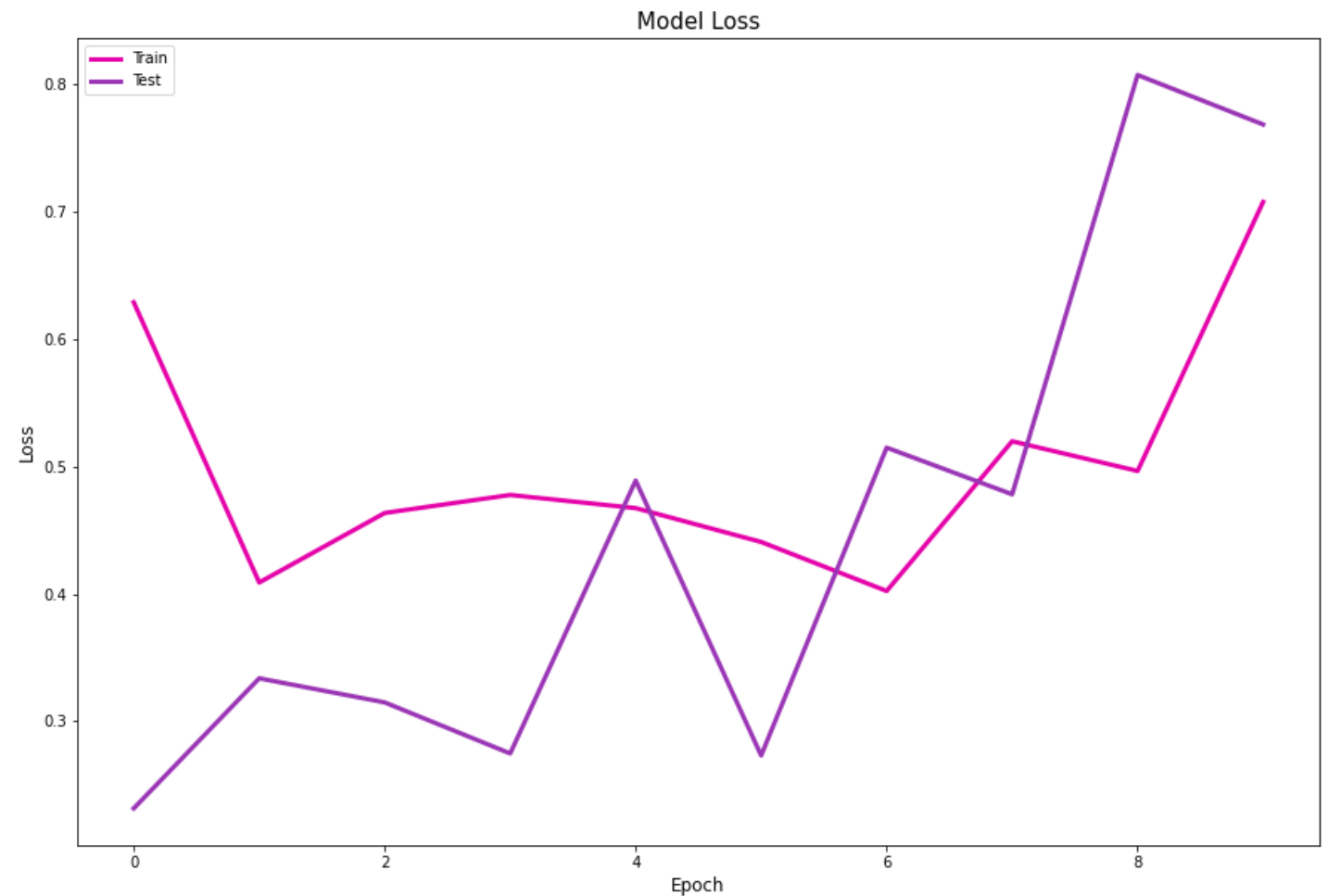
| | precision | recall | f1-score | support |
|-----------|-----------|--------|----------|---------|
| 0 | 0.00 | 0.00 | 0.00 | 116 |
| 1 | 0.98 | 0.54 | 0.70 | 317 |
| 2 | 0.76 | 1.00 | 0.87 | 855 |
| accuracy | | | 0.79 | 1288 |
| macro avg | 0.58 | 0.51 | 0.52 | 1288 |

weighted avg 0.75 0.79 0.75 1288

```
/usr/local/lib/python3.6/dist-packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))
```

In []:

```
plot_loss(history_6_3)
```



Base on the result, the weights in this model are not updating in a correct direction and it seems that this model will not reach to an optimal solution.

Using Adam Optimizer

In []:

```
adam_model = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.Adam())
adam_history = adam_model.fit(rescaled_train, validation_data=rescaled_test, epochs=10)
```

```
Epoch 1/10
161/161 [=====] - 144s 894ms/step - loss: 4.6470 - accuracy: 0.6933 - val_loss: 0.2442 - val_accuracy: 0.9092
Epoch 2/10
161/161 [=====] - 145s 901ms/step - loss: 0.3179 - accuracy: 0.8781 - val_loss: 0.2631 - val_accuracy: 0.8921
Epoch 3/10
161/161 [=====] - 144s 894ms/step - loss: 0.3083 - accuracy: 0.8892 - val_loss: 0.1883 - val_accuracy: 0.9332
Epoch 4/10
161/161 [=====] - 144s 894ms/step - loss: 0.2586 - accuracy: 0.9098 - val_loss: 0.2109 - val_accuracy: 0.9286
Epoch 5/10
161/161 [=====] - 144s 897ms/step - loss: 0.2854 - accuracy: 0.8989 - val_loss: 0.1798 - val_accuracy: 0.9387
Epoch 6/10
161/161 [=====] - 146s 911ms/step - loss: 0.2558 - accuracy: 0.8993 - val_loss: 0.1879 - val_accuracy: 0.9332
Epoch 7/10
161/161 [=====] - 144s 893ms/step - loss: 0.2283 - accuracy: 0.9151 - val_loss: 0.2217 - val_accuracy: 0.9130
Epoch 8/10
161/161 [=====] - 144s 894ms/step - loss: 0.2286 - accuracy: 0.9130 - val_loss: 0.2371 - val_accuracy: 0.9099
Epoch 9/10
161/161 [=====] - 143s 891ms/step - loss: 0.2155 - accuracy: 0.9146 - val_loss: 0.1804 - val_accuracy: 0.9441
Epoch 10/10
161/161 [=====] - 144s 894ms/step - loss: 0.2201 - accuracy: 0.9151 - val_loss: 0.2111 - val_accuracy: 0.9177
```

In []:

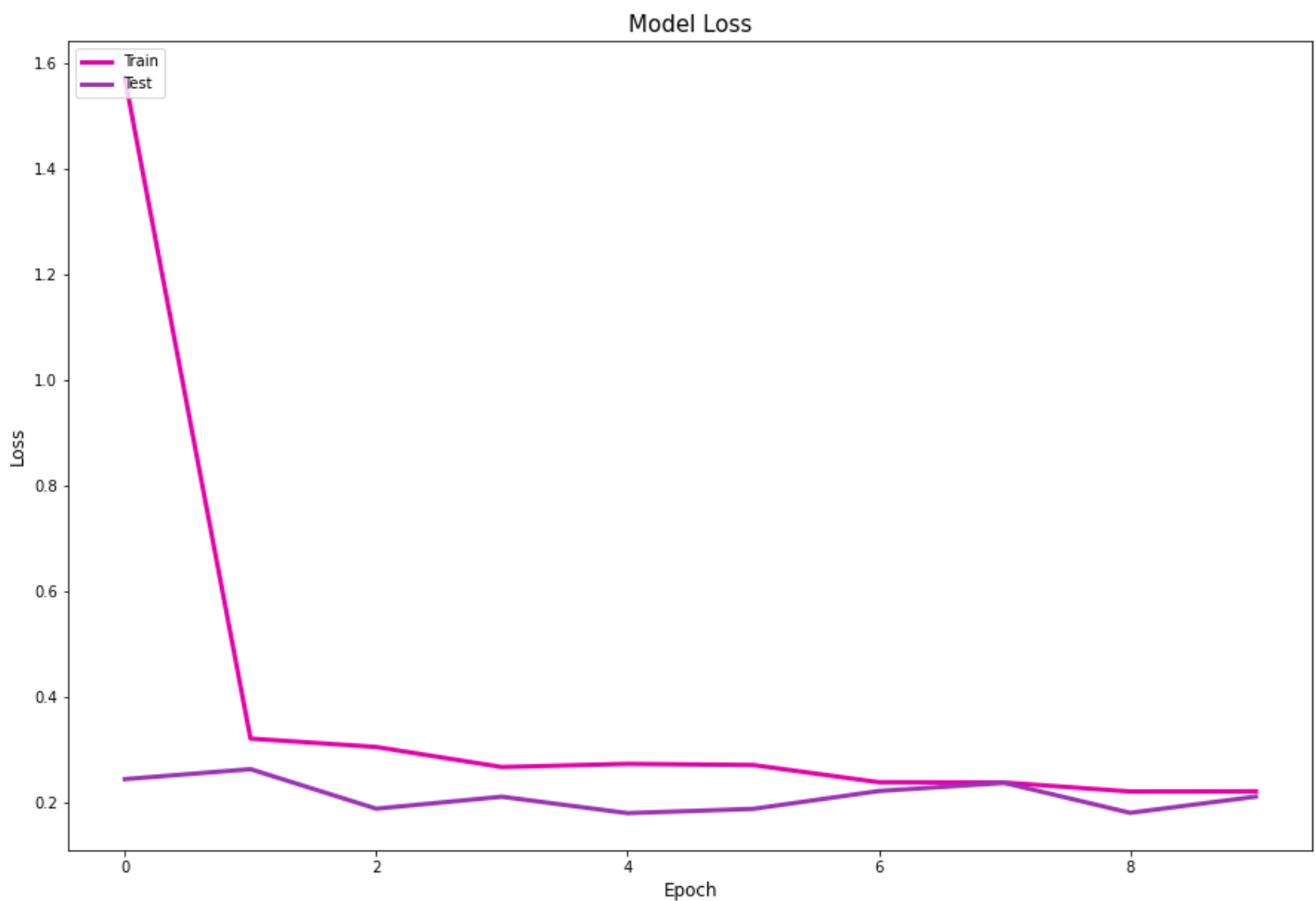
```
predict(adam_model, rescaled_test)
```

| | precision | recall | f1-score | support |
|---|-----------|--------|----------|---------|
| 0 | 0.89 | 0.97 | 0.93 | 116 |
| 1 | 0.80 | 0.94 | 0.86 | 317 |
| 2 | 0.98 | 0.90 | 0.94 | 855 |

| | accuracy | | 0.92 | 1288 |
|--------------|----------|------|------|------|
| macro avg | 0.89 | 0.94 | 0.91 | 1288 |
| weighted avg | 0.93 | 0.92 | 0.92 | 1288 |

In []:

```
plot_loss(adam_history)
```



Based on the result, Adam is much faster than SGD and the model converges in earlier epochs in comparison to SGD. Moreover, it works fine with default hyperparameters.

Question 7

In []:

```
model_7 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.Adam())  
history_7 = model_7.fit(rescaled_train, validation_data=rescaled_test, epochs=20)
```

Epoch 1/20

161/161 [=====] - 143s 888ms/step - loss: 2.9332 - accuracy: 0.6538 - val_loss: 0.3193 - val_accuracy: 0.8835

Epoch 2/20

161/161 [=====] - 142s 885ms/step - loss: 0.3843 - accuracy: 0.8498 - val_loss: 0.2422 - val_accuracy: 0.9115

Epoch 3/20

161/161 [=====] - 142s 885ms/step - loss: 0.3008 - accuracy: 0.8833 - val_loss: 0.2205 - val_accuracy: 0.9317

Epoch 4/20

161/161 [=====] - 142s 884ms/step - loss: 0.2527 - accuracy: 0.9028 - val_loss: 0.1813 - val_accuracy: 0.9387

Epoch 5/20
161/161 [=====] - 142s 883ms/step - loss: 0.2595 - accuracy: 0.9061 - val_loss: 0.2779 - val_accuracy: 0.8975
Epoch 6/20
161/161 [=====] - 142s 887ms/step - loss: 0.2534 - accuracy: 0.9093 - val_loss: 0.2232 - val_accuracy: 0.9138
Epoch 7/20
161/161 [=====] - 142s 886ms/step - loss: 0.2485 - accuracy: 0.9030 - val_loss: 0.3417 - val_accuracy: 0.8641
Epoch 8/20
161/161 [=====] - 142s 884ms/step - loss: 0.2548 - accuracy: 0.9072 - val_loss: 0.1790 - val_accuracy: 0.9410
Epoch 9/20
161/161 [=====] - 142s 885ms/step - loss: 0.2130 - accuracy: 0.9295 - val_loss: 0.2057 - val_accuracy: 0.9317
Epoch 10/20
161/161 [=====] - 143s 888ms/step - loss: 0.2231 - accuracy: 0.9175 - val_loss: 0.2318 - val_accuracy: 0.9123
Epoch 11/20
161/161 [=====] - 142s 883ms/step - loss: 0.2131 - accuracy: 0.9265 - val_loss: 0.1616 - val_accuracy: 0.9402
Epoch 12/20
161/161 [=====] - 143s 889ms/step - loss: 0.2147 - accuracy: 0.9210 - val_loss: 0.2123 - val_accuracy: 0.9348
Epoch 13/20
161/161 [=====] - 142s 884ms/step - loss: 0.2180 - accuracy: 0.9167 - val_loss: 0.1481 - val_accuracy: 0.9472
Epoch 14/20
161/161 [=====] - 142s 886ms/step - loss: 0.1964 - accuracy: 0.9260 - val_loss: 0.1781 - val_accuracy: 0.9309
Epoch 15/20
161/161 [=====] - 142s 884ms/step - loss: 0.1872 - accuracy: 0.9300 - val_loss: 0.1694 - val_accuracy: 0.9425
Epoch 16/20
161/161 [=====] - 145s 899ms/step - loss: 0.1813 - accuracy: 0.9381 - val_loss: 0.1603 - val_accuracy: 0.9418
Epoch 17/20
161/161 [=====] - 144s 896ms/step - loss: 0.1856 - accuracy: 0.9362 - val_loss: 0.1699 - val_accuracy: 0.9433
Epoch 18/20
161/161 [=====] - 143s 891ms/step - loss: 0.1714 - accuracy: 0.9450 - val_loss: 0.1454 - val_accuracy: 0.9464
Epoch 19/20
161/161 [=====] - 145s 899ms/step - loss: 0.1782 - accuracy: 0.9308 - val_loss: 0.1708 - val_accuracy: 0.9340
Epoch 20/20
161/161 [=====] - 144s 896ms/step - loss: 0.1954 - accuracy: 0.9222 - val_loss: 0.1659 - val_accuracy: 0.9441

In []:

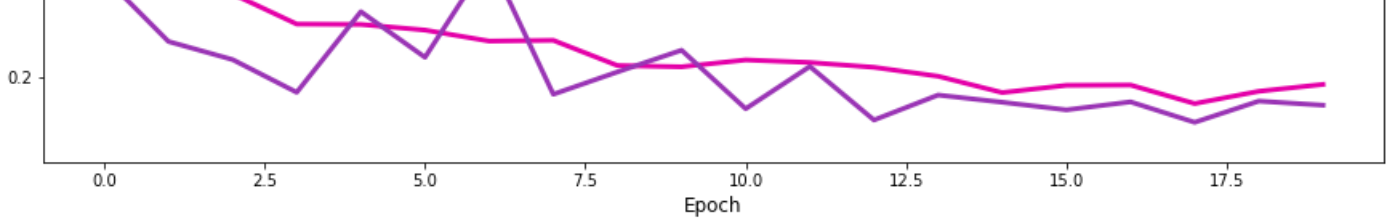
```
predict(model_7, rescaled_test)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.96 | 0.89 | 0.92 | 116 |
| 1 | 0.92 | 0.89 | 0.90 | 317 |
| 2 | 0.95 | 0.97 | 0.96 | 855 |
| accuracy | | | 0.94 | 1288 |
| macro avg | 0.94 | 0.92 | 0.93 | 1288 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1288 |

In []:

```
plot_loss(history_7)
```





Why we train neural network for multiple epochs?

In each epoch, weights are updated for several times and the whole data is checked. It is undeniable that we want to get good performance on non-training data and usually that takes more than one pass over the training data. Moreover, it is typical that gradient descent does not reach a global or local minimum after the first epoch. So, training just one epoch can lead to underfitting. As a result, we use multiple epochs.

Is it always good to train for more epochs?

If we set the number of epochs to a very large number, it may lead to have an overfitted model, because we train the data for many times and our model will start to model the noises in our data and it is a sign of overfitting and the accuracy of the test data will be reduced.

Techniques to avoid overfitting by controlling the number of epochs:

- **Early stopping:** in this case we set the number of epochs to a really high number and you turn off the training when the improvement over next epochs is not satisfying.
- **Model Checkpoint:** here we once again set up a really high number of epochs and we simply save only the best model to a metric chosen. Once again we have a special callback for this scenario.

Question 8

Why MSE loss function is not good for classification problems?

There are two reasons that MSE loss function is not good for classification problems:

Firstly, when we use MSE, it means we assume the underlying data has been generated from a normal distribution. In Bayesian terms this means we assume a Gaussian prior. While in fact, our dataset can be classified into three categories. As a result, it is not from a normal distribution.

Secondly, the MSE function is non-convex for classification. In other words, if a classification model is trained with MSE loss function, it is not guaranteed to minimize the loss function. This is because MSE function expects real-valued inputs in range $(-\infty, \infty)$, while classification models output probabilities in range $(0,1)$ through the activation function.

In which problems we use MSE loss function?

We use this loss function for regression problems.

In []:

```
model_8 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.Adam(), loss=losses.mse)
history_8 = model_8.fit(rescaled_train, validation_data=rescaled_test, epochs=20)
```

```
Epoch 1/20
161/161 [=====] - 143s 888ms/step - loss: 0.2290 - accuracy: 0.6533 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 2/20
161/161 [=====] - 142s 883ms/step - loss: 0.2254 - accuracy: 0.6619 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 3/20
161/161 [=====] - 142s 885ms/step - loss: 0.2205 - accuracy: 0.6692 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 4/20
161/161 [=====] - 142s 883ms/step - loss: 0.2235 - accuracy: 0.6648 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 5/20
161/161 [=====] - 142s 884ms/step - loss: 0.2322 - accuracy: 0.6517 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 6/20
161/161 [=====] - 144s 893ms/step - loss: 0.2231 - accuracy: 0.6654 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 7/20
161/161 [=====] - 142s 883ms/step - loss: 0.2283 - accuracy: 0.6575 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 8/20
161/161 [=====] - 141s 877ms/step - loss: 0.2266 - accuracy: 0.6601 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 9/20
161/161 [=====] - 141s 874ms/step - loss: 0.2236 - accuracy: 0.6646 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 10/20
161/161 [=====] - 143s 889ms/step - loss: 0.2205 - accuracy: 0.6693 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 11/20
161/161 [=====] - 141s 878ms/step - loss: 0.2184 - accuracy: 0.6724 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 12/20
161/161 [=====] - 141s 878ms/step - loss: 0.2184 - accuracy: 0.6724 - val_loss: 0.2241 - val_accuracy: 0.6638
```


Epoch 12/20
161/161 [=====] - 142s 885ms/step - loss: 0.2252 - accuracy: 0.6622 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 13/20
161/161 [=====] - 143s 887ms/step - loss: 0.2171 - accuracy: 0.6743 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 14/20
161/161 [=====] - 144s 896ms/step - loss: 0.2240 - accuracy: 0.6640 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 15/20
161/161 [=====] - 150s 935ms/step - loss: 0.2225 - accuracy: 0.6663 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 16/20
161/161 [=====] - 147s 916ms/step - loss: 0.2343 - accuracy: 0.6485 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 17/20
161/161 [=====] - 148s 920ms/step - loss: 0.2237 - accuracy: 0.6645 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 18/20
161/161 [=====] - 152s 944ms/step - loss: 0.2267 - accuracy: 0.6600 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 19/20
161/161 [=====] - 151s 937ms/step - loss: 0.2243 - accuracy: 0.6636 - val_loss: 0.2241 - val_accuracy: 0.6638
Epoch 20/20
161/161 [=====] - 150s 933ms/step - loss: 0.2269 - accuracy: 0.6597 - val_loss: 0.2241 - val_accuracy: 0.6638

In []:

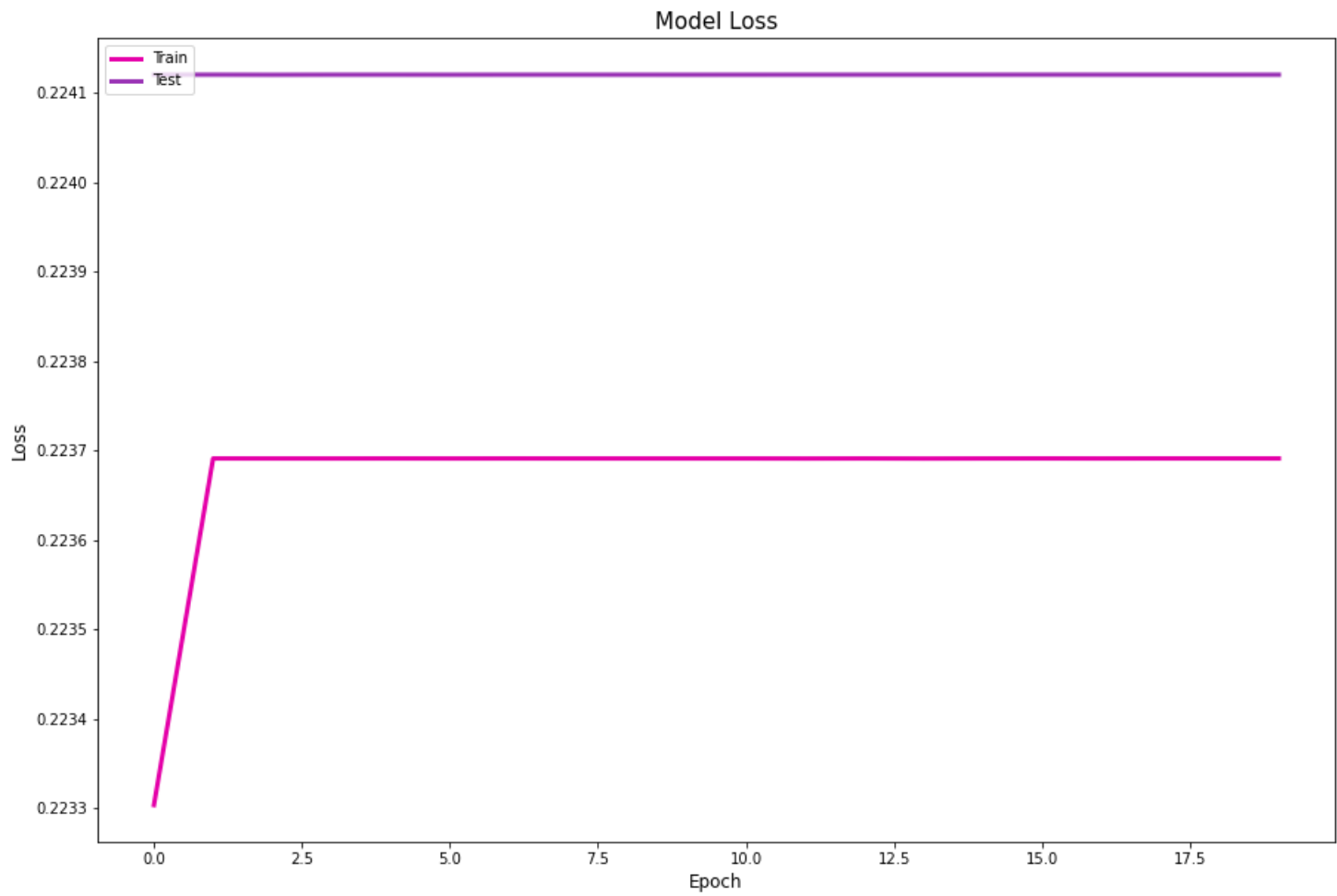
```
predict(model_8, rescaled_test)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.00 | 0.00 | 0.00 | 116 |
| 1 | 0.00 | 0.00 | 0.00 | 317 |
| 2 | 0.66 | 1.00 | 0.80 | 855 |
| accuracy | | | 0.66 | 1288 |
| macro avg | 0.22 | 0.33 | 0.27 | 1288 |
| weighted avg | 0.44 | 0.66 | 0.53 | 1288 |

/usr/local/lib/python3.6/dist-packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))

In []:

```
plot_loss(history_8)
```



Based on the result, by using this loss function, the rate of the learning process decreased and it seems that it will not converge and it refers to the reasons that we mentioned above.

Question 9

As we mentioned in previous parts, if we train our model for more epochs, the risk of overfitting will be increased. One of the techniques to avoid overfitting is to use regularization methods that we want to discuss in this part.

Using Regularization

In []:

```
model_9_1 = make_feedforward_network(4800, 3200, 1600, optimizer=optimizers.Adam(), regularizer=regularizers.l2(0.0001))
history_9_1 = model_9_1.fit(rescaled_train, validation_data=rescaled_test, epochs=20)
```

Epoch 1/20
161/161 [=====] - 153s 946ms/step - loss: 4.4287 - accuracy: 0.6691 - val_loss: 0.8305 - val_accuracy: 0.9115
Epoch 2/20
161/161 [=====] - 151s 939ms/step - loss: 0.8755 - accuracy: 0.8771 - val_loss: 0.6443 - val_accuracy: 0.9325
Epoch 3/20
161/161 [=====] - 151s 939ms/step - loss: 0.7080 - accuracy: 0.8996 - val_loss: 0.6388 - val_accuracy: 0.8882
Epoch 4/20
161/161 [=====] - 152s 945ms/step - loss: 0.6416 - accuracy: 0.8865 - val_loss: 0.6081 - val_accuracy: 0.8905
Epoch 5/20
161/161 [=====] - 150s 935ms/step - loss: 0.5704 - accuracy: 0.8912 - val_loss: 0.4230 - val_accuracy: 0.9402
Epoch 6/20
161/161 [=====] - 151s 939ms/step - loss: 0.4804 - accuracy: 0.9126 - val_loss: 0.4557 - val_accuracy: 0.9130
Epoch 7/20
161/161 [=====] - 150s 934ms/step - loss: 0.4556 - accuracy: 0.9036 - val_loss: 0.4018 - val_accuracy: 0.9224
Epoch 8/20
161/161 [=====] - 149s 927ms/step - loss: 0.4291 - accuracy: 0.9028 - val_loss: 0.4138 - val_accuracy: 0.8967
Epoch 9/20
161/161 [=====] - 150s 931ms/step - loss: 0.4038 - accuracy: 0.9045 - val_loss: 0.3299 - val_accuracy: 0.9449
Epoch 10/20
161/161 [=====] - 150s 932ms/step - loss: 0.3852 - accuracy: 0.9111 - val_loss: 0.3215 - val_accuracy: 0.9387
Epoch 11/20
161/161 [=====] - 150s 929ms/step - loss: 0.3318 - accuracy: 0.9246 - val_loss: 0.3555 - val_accuracy: 0.8983
Epoch 12/20
161/161 [=====] - 151s 937ms/step - loss: 0.3689 - accuracy: 0.9023 - val_loss: 0.3065 - val_accuracy: 0.9309
Epoch 13/20
161/161 [=====] - 151s 942ms/step - loss: 0.3239 - accuracy: 0.9224 - val_loss: 0.2546 - val_accuracy: 0.9433
Epoch 14/20
161/161 [=====] - 151s 940ms/step - loss: 0.3729 - accuracy: 0.8958 - val_loss: 0.2583 - val_accuracy: 0.9441
Epoch 15/20
161/161 [=====] - 148s 919ms/step - loss: 0.3075 - accuracy: 0.9224 - val_loss: 0.2844 - val_accuracy: 0.9193
Epoch 16/20
161/161 [=====] - 146s 909ms/step - loss: 0.2678 - accuracy: 0.9320 - val_loss: 0.2456 - val_accuracy: 0.9402
Epoch 17/20
161/161 [=====] - 147s 913ms/step - loss: 0.2560 - accuracy: 0.9329 - val_loss: 0.3563 - val_accuracy: 0.8960
Epoch 18/20
161/161 [=====] - 144s 895ms/step - loss: 0.2911 - accuracy: 0.9168 - val_loss: 0.2288 - val_accuracy: 0.9457
Epoch 19/20
161/161 [=====] - 142s 883ms/step - loss: 0.2398 - accuracy: 0.9311 - val_loss: 0.4056 - val_accuracy: 0.8719
Epoch 20/20
161/161 [=====] - 142s 883ms/step - loss: 0.2965 - accuracy: 0.9047 - val_loss: 0.2569 - val_accuracy: 0.9278

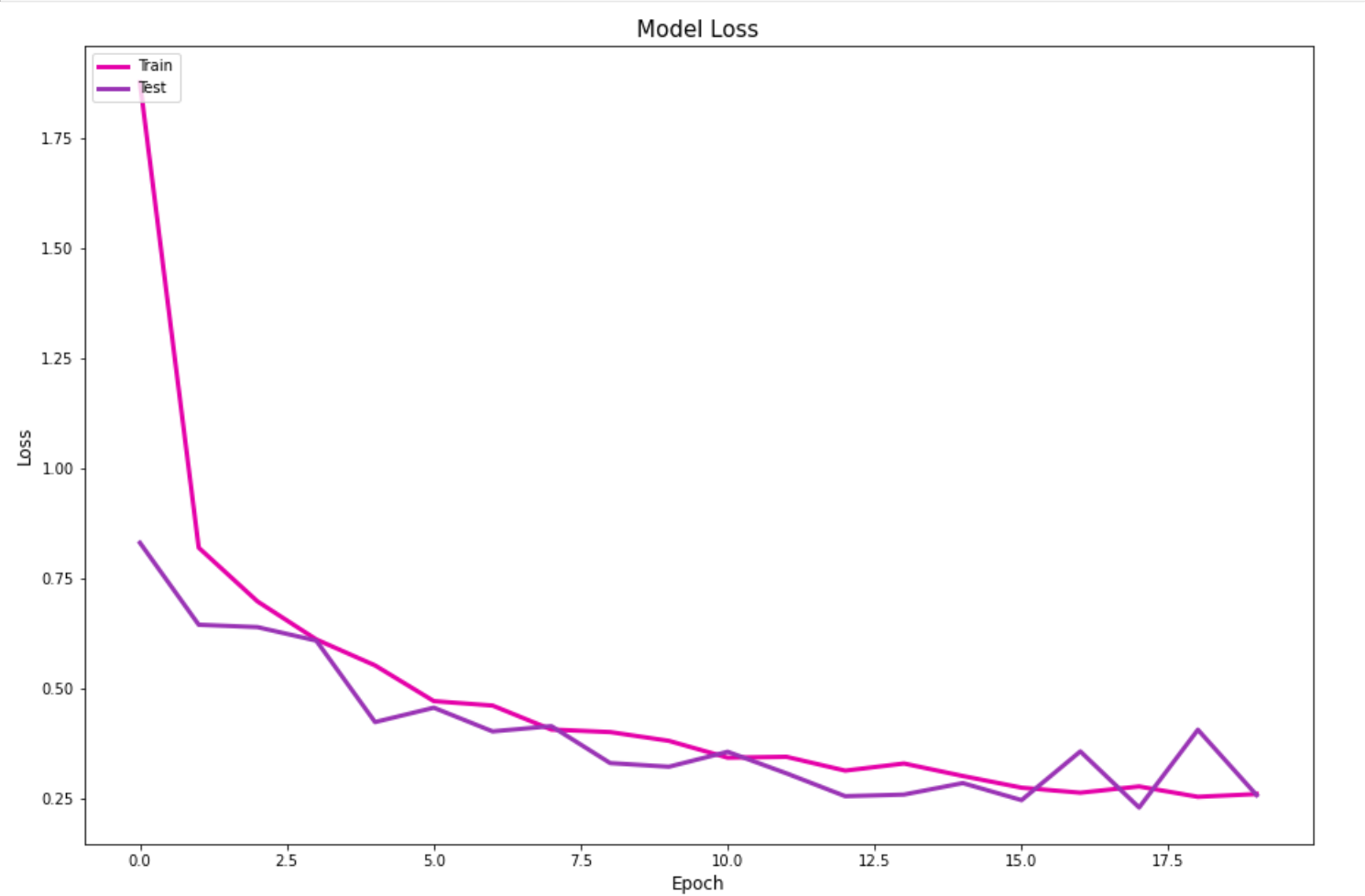
In []:

```
predict(model_9_1, rescaled_test)
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.79 | 0.97 | 0.87 | 116 |
| 1 | 0.95 | 0.84 | 0.89 | 317 |
| 2 | 0.94 | 0.95 | 0.95 | 855 |
| accuracy | | | 0.93 | 1288 |
| macro avg | 0.89 | 0.92 | 0.90 | 1288 |
| weighted avg | 0.93 | 0.93 | 0.93 | 1288 |

In []:

```
plot_loss(history_9_1)
```



Based on the result, the loss curve is smoother and it shows that our model has not faced with overfitting, during the learning process.

Using Dropout Layers

```
In []:
model_9_2 = make_feedforward_network_with_dropout(4800, 3200, 1600, 0.1, optimizer=optimizers.Adam())
history_9_2 = model_9_2.fit(rescaled_train, validation_data=rescaled_test, epochs=20)
```

Epoch 1/20
161/161 [=====] - 144s 893ms/step - loss: 3.4712 - accuracy: 0.6859 - val_loss: 0.3228 - val_accuracy: 0.8866
Epoch 2/20
161/161 [=====] - 143s 887ms/step - loss: 0.3473 - accuracy: 0.8732 - val_loss: 0.2979 - val_accuracy: 0.9006
Epoch 3/20
161/161 [=====] - 142s 884ms/step - loss: 0.3264 - accuracy: 0.8768 - val_loss: 0.2264 - val_accuracy: 0.9084
Epoch 4/20
161/161 [=====] - 142s 883ms/step - loss: 0.3143 - accuracy: 0.8805 - val_loss: 0.2353 - val_accuracy: 0.9068
Epoch 5/20
161/161 [=====] - 142s 880ms/step - loss: 0.2727 - accuracy: 0.8971 - val_loss: 0.2000 - val_accuracy: 0.9293
Epoch 6/20
161/161 [=====] - 142s 885ms/step - loss: 0.2600 - accuracy: 0.9051 - val_loss: 0.2038 - val_accuracy: 0.9224
Epoch 7/20
161/161 [=====] - 143s 888ms/step - loss: 0.2623 - accuracy: 0.9043 - val_loss: 0.2123 - val_accuracy: 0.9224
Epoch 8/20
161/161 [=====] - 142s 886ms/step - loss: 0.2751 - accuracy: 0.9030 - val_loss: 0.1953 - val_accuracy: 0.9293
Epoch 9/20
161/161 [=====] - 143s 887ms/step - loss: 0.2432 - accuracy: 0.9081 - val_loss: 0.1711 - val_accuracy: 0.9379
Epoch 10/20
161/161 [=====] - 142s 886ms/step - loss: 0.2258 - accuracy: 0.9119 - val_loss: 0.1955 - val_accuracy: 0.9356
Epoch 11/20
161/161 [=====] - 142s 887ms/step - loss: 0.2328 - accuracy: 0.9183 - val_loss: 0.2244 - val_accuracy: 0.9239
Epoch 12/20
161/161 [=====] - 143s 886ms/step - loss: 0.2529 - accuracy: 0.9031 - val_loss: 0.1892 - val_accuracy: 0.9332
Epoch 13/20
161/161 [=====] - 143s 887ms/step - loss: 0.2287 - accuracy: 0.9137 - val_loss: 0.2096 - val_accuracy: 0.9394
Epoch 14/20
161/161 [=====] - 142s 885ms/step - loss: 0.2343 - accuracy: 0.9087 - val_loss: 0.1663 - val_accuracy: 0.9449
Epoch 15/20
161/161 [=====] - 142s 885ms/step - loss: 0.2347 - accuracy: 0.9181 - val_loss: 0.1776 - val_accuracy: 0.9332
Epoch 16/20
161/161 [=====] - 142s 886ms/step - loss: 0.2268 - accuracy: 0.9165 - val_loss: 0.1740 - val_accuracy: 0.9441

Epoch 17/20
161/161 [=====] - 142s 885ms/step - loss: 0.2233 - accuracy: 0.9185 - val_loss: 0.1671 - val_accuracy: 0.9433
Epoch 18/20
161/161 [=====] - 142s 884ms/step - loss: 0.2483 - accuracy: 0.9040 - val_loss: 0.2046 - val_accuracy: 0.9247
Epoch 19/20
161/161 [=====] - 142s 883ms/step - loss: 0.2187 - accuracy: 0.9191 - val_loss: 0.1807 - val_accuracy: 0.9356
Epoch 20/20
161/161 [=====] - 142s 885ms/step - loss: 0.2055 - accuracy: 0.9253 - val_loss: 0.1659 - val_accuracy: 0.9449

In []:

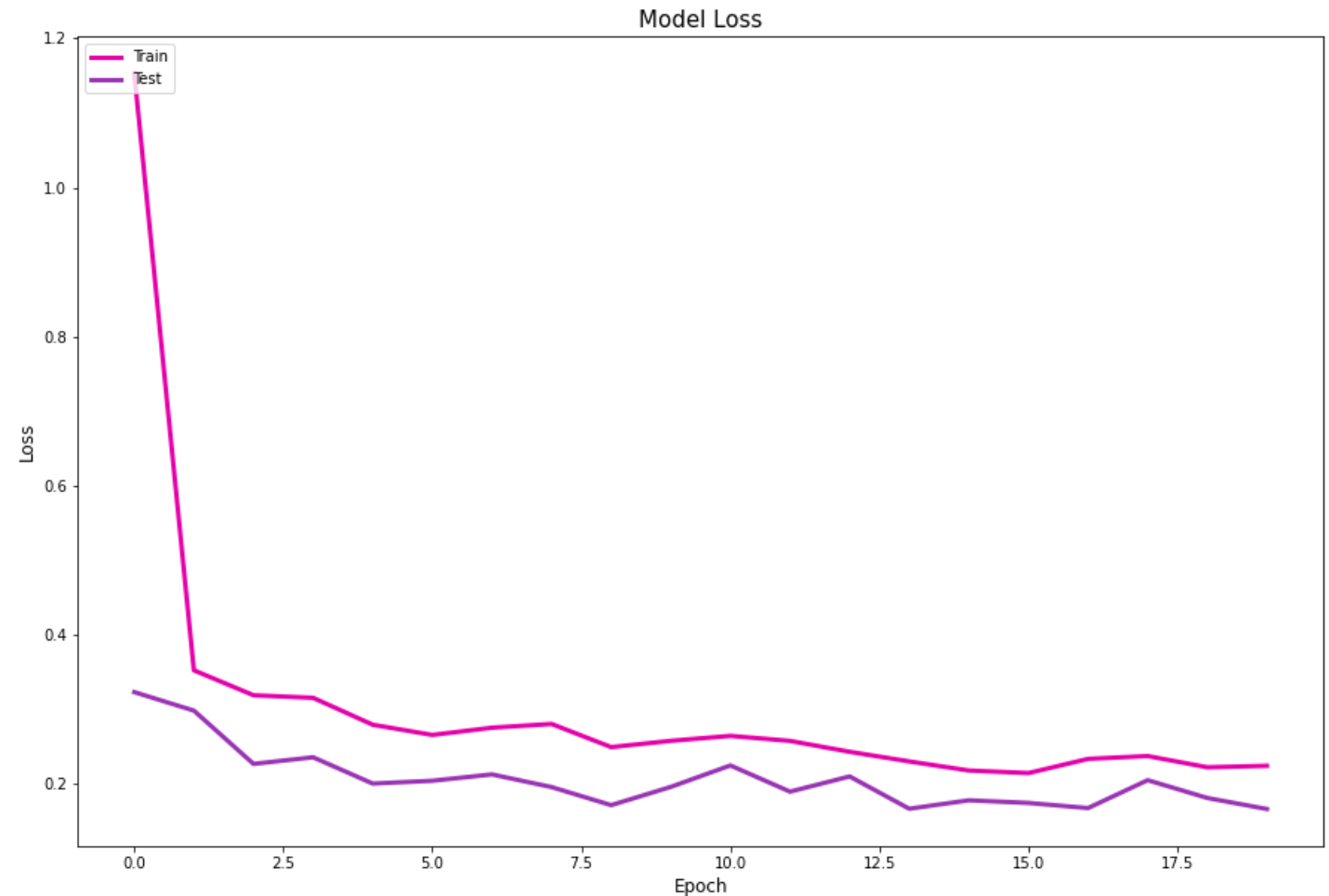
```
predict(model_9_2, rescaled_test)
```

| | precision | recall | f1-score | support |
|---|-----------|--------|----------|---------|
| 0 | 0.95 | 0.91 | 0.93 | 116 |
| 1 | 0.89 | 0.92 | 0.91 | 317 |
| 2 | 0.96 | 0.96 | 0.96 | 855 |

| | accuracy | | 0.94 | 1288 |
|--------------|----------|------|------|------|
| macro avg | 0.93 | 0.93 | 0.93 | 1288 |
| weighted avg | 0.95 | 0.94 | 0.95 | 1288 |

In []:

```
plot_loss(history_9_2)
```



Here, the loss curve is smoother than the previous part and it seems that this method can deal with overfitting in a better way.

Conclusion

In this computer assignment we learned that neural networks are good methods to solve image classification problems. Moreover, we were introduced to the method of Keras library. Finally, we learned how to tune the parameters in order to reach an optimal model.