README.md

Deeper Neural Networks - Lab

Introduction

In this lesson, we'll dig deeper into the work horse of deep learning, ***Multi-Layer Perceptrons***! We'll build and train a couple of different MLPs with Keras and explore the tradeoffs that come with adding extra hidden layers. We'll also try switching between some of the activation functions we learned about in the previous lesson to see how they affect training and performance.

Objectives

- Build a deep neural network using Keras

Getting Started

Run the cell below to import everything we'll need for this lab.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import keras
from keras.models import Sequential
from keras.layers import Dense
from sklearn.datasets import load_breast_cancer
from sklearn.preprocessing import StandardScaler, LabelBinarizer
```

For this lab, we'll be working with the Boston Breast Cancer Dataset. Although we're importing this dataset directly from scikit-learn, the Kaggle link above contains a detailed explanation of the dataset, in case you're interested. We recommend you take a minute to familiarize yourself with the dataset before digging in.

In the cell below:

- Call `load_breast_cancer()` to store the dataset
- Access the `.data` , `.target` , and `.feature_names` attributes and store them in the appropriate variables below

```
bc_dataset = load_breast_cancer()
data = bc_dataset.data
target = bc_dataset.target
col_names = bc_dataset.feature_names
```

Now, let's create a DataFrame so that we can see the data and explore it a bit more easily with the column names attached.

- In the cell below, create a pandas DataFrame from `data` (use `col_names` for column names)
- Print the `.head()` of the DataFrame

```
df = pd.DataFrame(data, columns=col_names)
df.head()
```

<style scoped> .dataframe tbody tr th:only-of-type { vertical-align: middle; }

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

</style>

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	me conc
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.300
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.086
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.197
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.241
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.198

5 rows × 30 columns

Getting the Data Ready for Deep Learning

In order to pass this data into a neural network, we'll need to make sure that the data:

- is purely numerical
- contains no missing values
- is normalized

Let's begin by calling the DataFrame's `.info()` method to check the datatype of each feature.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 30 columns):
```

```
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   mean radius              569 non-null    float64
 1   mean texture             569 non-null    float64
 2   mean perimeter           569 non-null    float64
 3   mean area                569 non-null    float64
 4   mean smoothness          569 non-null    float64
 5   mean compactness         569 non-null    float64
 6   mean concavity           569 non-null    float64
 7   mean concave points      569 non-null    float64
 8   mean symmetry            569 non-null    float64
 9   mean fractal dimension   569 non-null    float64
 10  radius error             569 non-null    float64
 11  texture error            569 non-null    float64
 12  perimeter error          569 non-null    float64
 13  area error               569 non-null    float64
 14  smoothness error         569 non-null    float64
 15  compactness error        569 non-null    float64
 16  concavity error          569 non-null    float64
 17  concave points error     569 non-null    float64
 18  symmetry error           569 non-null    float64
 19  fractal dimension error  569 non-null    float64
 20  worst radius             569 non-null    float64
 21  worst texture            569 non-null    float64
 22  worst perimeter          569 non-null    float64
 23  worst area               569 non-null    float64
 24  worst smoothness         569 non-null    float64
 25  worst compactness        569 non-null    float64
 26  worst concavity          569 non-null    float64
 27  worst concave points     569 non-null    float64
 28  worst symmetry           569 non-null    float64
 29  worst fractal dimension  569 non-null    float64
dtypes: float64(30)
memory usage: 133.5 KB
```

From the output above, we can see that the entire dataset is already in numerical format. We can also see from the counts that each feature has the same number of entries as the number of rows in the DataFrame -- that means that no feature contains any missing values. Great!

Now, let's check to see if our data needs to be normalized. Instead of doing statistical tests here, let's just take a quick look at the `.head()` of the DataFrame again. Do this in the cell below.

```
df.head()
```

<style scoped> .dataframe tbody tr th:only-of-type { vertical-align: middle; }

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

</style>

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	me conc
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.300
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.086
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.197
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.241
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.198

5 rows × 30 columns

As we can see from comparing `mean radius` and `mean area` , columns are clearly on different scales, which means that we need to normalize our dataset. To do this, we'll make use of scikit-learn's `StandardScaler()` class.

In the cell below, instantiate a `StandardScaler` and use it to create a normalized version of our dataset.

```
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data)
```

Binarizing our Labels

If you took a look at the data dictionary on Kaggle, then you probably noticed the target for this dataset is to predict if the sample is "M" (Malignant) or "B" (Benign). This means that this is a ***Binary Classification*** task, so we'll need to binarize our labels.

In the cell below, make use of scikit-learn's `LabelBinarizer()` class to create a binarized version of our labels.

```
binarizer = LabelBinarizer()
labels = binarizer.fit_transform(target)
```

Building our MLP

Now, we'll build a small ***Multi-Layer Perceptron*** using Keras in the cell below. Our first model will act as a baseline, and then we'll make it bigger to see what happens to model performance.

In the cell below:

- Instantiate a `Sequential()` Keras model
- Use the model's `.add()` method to add a `Dense` layer with 10 neurons and a `'tanh'` activation function. Also set the `input_shape` attribute to `(30,)`, since we have 30 features
- Since this is a binary classification task, the output layer should be a `Dense` layer with a single neuron, and the activation set to `'sigmoid'`

```
model_1 = Sequential()
model_1.add(Dense(5, activation='tanh', input_shape=(30,)))
model_1.add(Dense(1, activation='sigmoid'))
```

Compiling the Model

Now that we've created the model, the next step is to compile it.

In the cell below, compile the model. Set the following hyperparameters:

- `loss='binary_crossentropy'`
- `optimizer='sgd'`
- `metrics=['acc']`

```
model_1.compile(loss='binary_crossentropy', optimizer='sgd', metrics=['acc'])
```

Fitting the Model

Now, let's fit the model. Set the following hyperparameters:

- `epochs=25`
- `batch_size=1`
- `validation_split=0.2`

```python
results_1 = model_1.fit(scaled_data, labels, epochs=25, batch_size=1, validation_spl
```

```
Epoch 1/25
455/455 [==============================] - 0s 661us/step - loss: 0.2576 - acc:
0.9275 - val_loss: 0.2038 - val_acc: 0.9298
Epoch 2/25
455/455 [==============================] - 0s 459us/step - loss: 0.1198 - acc:
0.9692 - val_loss: 0.1543 - val_acc: 0.9298
Epoch 3/25
455/455 [==============================] - 0s 438us/step - loss: 0.0925 - acc:
0.9692 - val_loss: 0.1336 - val_acc: 0.9386
Epoch 4/25
455/455 [==============================] - 0s 436us/step - loss: 0.0811 - acc:
0.9758 - val_loss: 0.1254 - val_acc: 0.9386
Epoch 5/25
455/455 [==============================] - 0s 429us/step - loss: 0.0739 - acc:
0.9758 - val_loss: 0.1128 - val_acc: 0.9561
Epoch 6/25
455/455 [==============================] - 0s 430us/step - loss: 0.0698 - acc:
0.9780 - val_loss: 0.1128 - val_acc: 0.9561
Epoch 7/25
455/455 [==============================] - 0s 431us/step - loss: 0.0664 - acc:
0.9780 - val_loss: 0.1118 - val_acc: 0.9561
Epoch 8/25
455/455 [==============================] - 0s 443us/step - loss: 0.0635 - acc:
0.9758 - val_loss: 0.1032 - val_acc: 0.9649
Epoch 9/25
455/455 [==============================] - 0s 439us/step - loss: 0.0618 - acc:
0.9802 - val_loss: 0.1032 - val_acc: 0.9561
Epoch 10/25
455/455 [==============================] - 0s 436us/step - loss: 0.0598 - acc:
0.9780 - val_loss: 0.0976 - val_acc: 0.9649
Epoch 11/25
455/455 [==============================] - 0s 428us/step - loss: 0.0583 - acc:
0.9780 - val_loss: 0.0988 - val_acc: 0.9561
Epoch 12/25
455/455 [==============================] - 0s 429us/step - loss: 0.0567 - acc:
0.9802 - val_loss: 0.1027 - val_acc: 0.9561
Epoch 13/25
```

```
455/455 [==============================] - 0s 414us/step - loss: 0.0554 - acc:
0.9802 - val_loss: 0.1018 - val_acc: 0.9561
Epoch 14/25
455/455 [==============================] - 0s 425us/step - loss: 0.0537 - acc:
0.9824 - val_loss: 0.0952 - val_acc: 0.9649
Epoch 15/25
455/455 [==============================] - 0s 430us/step - loss: 0.0525 - acc:
0.9824 - val_loss: 0.0947 - val_acc: 0.9649
Epoch 16/25
455/455 [==============================] - 0s 428us/step - loss: 0.0515 - acc:
0.9802 - val_loss: 0.0955 - val_acc: 0.9649
Epoch 17/25
455/455 [==============================] - 0s 414us/step - loss: 0.0499 - acc:
0.9802 - val_loss: 0.1032 - val_acc: 0.9561
Epoch 18/25
455/455 [==============================] - 0s 428us/step - loss: 0.0487 - acc:
0.9802 - val_loss: 0.0979 - val_acc: 0.9649
Epoch 19/25
455/455 [==============================] - 0s 441us/step - loss: 0.0469 - acc:
0.9824 - val_loss: 0.0928 - val_acc: 0.9737
Epoch 20/25
455/455 [==============================] - 0s 414us/step - loss: 0.0471 - acc:
0.9802 - val_loss: 0.0964 - val_acc: 0.9649
Epoch 21/25
455/455 [==============================] - 0s 434us/step - loss: 0.0455 - acc:
0.9802 - val_loss: 0.0943 - val_acc: 0.9649
Epoch 22/25
455/455 [==============================] - 0s 439us/step - loss: 0.0453 - acc:
0.9846 - val_loss: 0.0964 - val_acc: 0.9649
Epoch 23/25
455/455 [==============================] - 0s 432us/step - loss: 0.0440 - acc:
0.9846 - val_loss: 0.0948 - val_acc: 0.9561
Epoch 24/25
455/455 [==============================] - 0s 430us/step - loss: 0.0432 - acc:
0.9846 - val_loss: 0.0948 - val_acc: 0.9561
Epoch 25/25
455/455 [==============================] - 0s 428us/step - loss: 0.0427 - acc:
0.9868 - val_loss: 0.1025 - val_acc: 0.9561
```

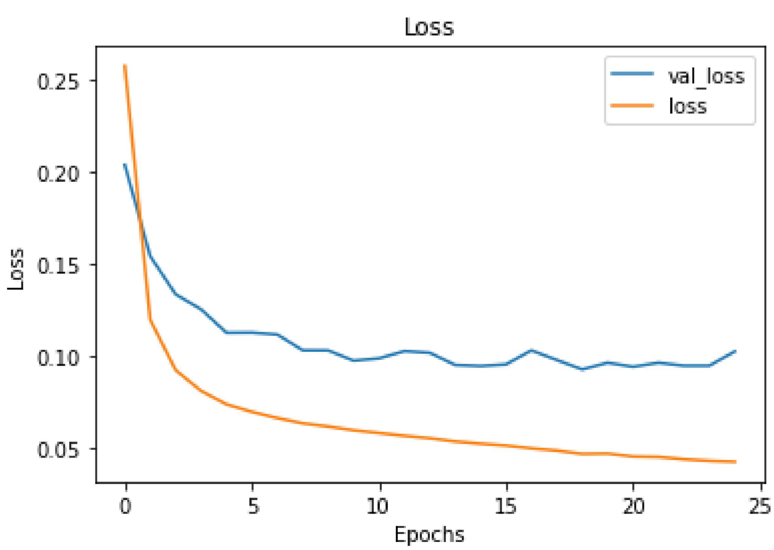
Note that when you call a Keras model's `.fit()` method, it returns a Keras callback containing information on the training process of the model. If you examine the callback's `.history` attribute, you'll find a dictionary containing both the training and validation loss, as well as any metrics we specified when compiling the model (in this case, just accuracy).

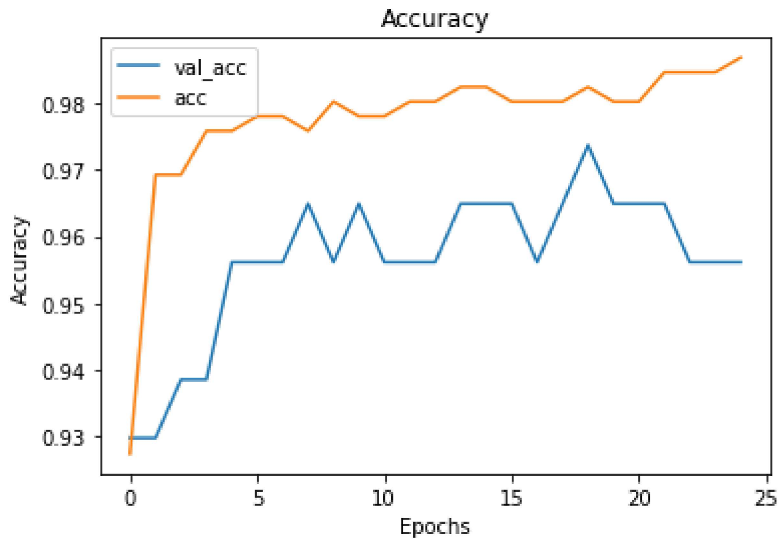
Let's quickly plot our validation and accuracy curves and see if we notice anything. Since we'll want to do this anytime we train an MLP, its worth wrapping this code in a function so that we can easily reuse it.

In the cell below, we created a function for visualizing the loss and accuracy metrics.

```
def visualize_training_results(results):
    history = results.history
    plt.figure()
    plt.plot(history['val_loss'])
    plt.plot(history['loss'])
    plt.legend(['val_loss', 'loss'])
    plt.title('Loss')
    plt.xlabel('Epochs')
    plt.ylabel('Loss')
    plt.show()

    plt.figure()
    plt.plot(history['val_acc'])
    plt.plot(history['acc'])
    plt.legend(['val_acc', 'acc'])
    plt.title('Accuracy')
    plt.xlabel('Epochs')
    plt.ylabel('Accuracy')
    plt.show()
```

```
visualize_training_results(results_1)
```





Detecting Overfitting

You'll probably notice that the model did pretty well! It's always recommended to visualize your training and validation metrics against each other after training a model. By plotting them like this, we can easily detect when the model is starting to overfit. We can tell that this is happening by seeing the model's training performance steadily improve long after the validation performance plateaus. We can see that in the plots above as the training loss continues to decrease and the training accuracy continues to increase, and the distance between the two lines gets greater as the epochs gets higher.

Iterating on the Model

By adding another hidden layer, we can a given the model the ability to capture more high-level abstraction in the data. However, increasing the depth of the model also increases the amount of data the model needs to converge to answer, because with a more complex model comes the "Curse of Dimensionality", thanks to all the extra trainable parameters that come from adding more size to our network.

If there is complexity in the data that our smaller model was not big enough to catch, then a larger model may improve performance. However, if our dataset isn't big enough for the new, larger model, then we may see performance decrease as then model "thrashes" about a bit, failing to converge. Let's try and see what happens.

In the cell below, recreate the model that you created above, with one exception. In the model below, add a second `Dense` layer with `'tanh'` activation function and 5 neurons after the first. The network's output layer should still be a `Dense` layer with a single neuron and a `'sigmoid'` activation function, since this is still a binary classification task.

Create, compile, and fit the model in the cells below, and then visualize the results to compare the history.

```
model_2 = Sequential()
model_2.add(Dense(10, activation='tanh', input_shape=(30,)))
model_2.add(Dense(5, activation='tanh'))
model_2.add(Dense(1, activation='sigmoid'))
```

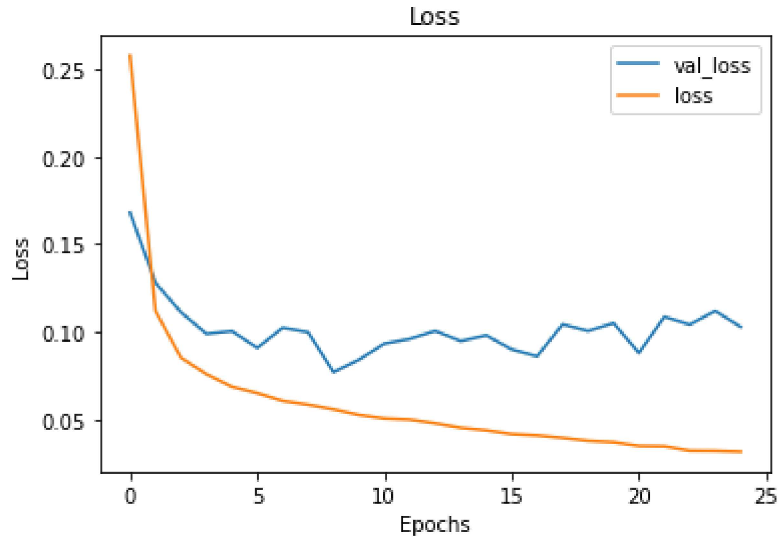
```
model_2.compile(loss='binary_crossentropy', optimizer='sgd', metrics=['acc'])
```

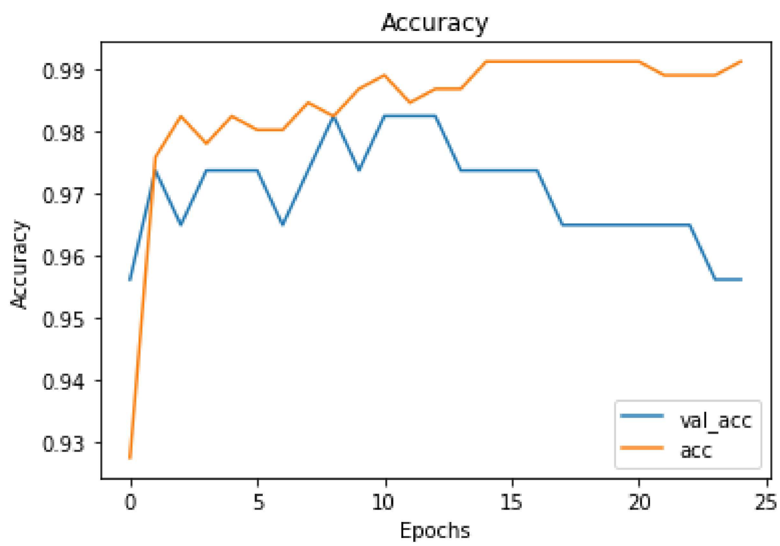
```
results_2 = model_2.fit(scaled_data, labels, epochs=25, batch_size=1, validation_spl
```

```
Epoch 1/25
455/455 [==============================] - 0s 616us/step - loss: 0.2577 - acc:
0.9275 - val_loss: 0.1680 - val_acc: 0.9561
Epoch 2/25
455/455 [==============================] - 0s 453us/step - loss: 0.1120 - acc:
0.9758 - val_loss: 0.1281 - val_acc: 0.9737
Epoch 3/25
455/455 [==============================] - 0s 460us/step - loss: 0.0854 - acc:
0.9824 - val_loss: 0.1113 - val_acc: 0.9649
Epoch 4/25
455/455 [==============================] - 0s 451us/step - loss: 0.0760 - acc:
0.9780 - val_loss: 0.0990 - val_acc: 0.9737
Epoch 5/25
455/455 [==============================] - 0s 449us/step - loss: 0.0688 - acc:
0.9824 - val_loss: 0.1006 - val_acc: 0.9737
Epoch 6/25
455/455 [==============================] - 0s 452us/step - loss: 0.0651 - acc:
0.9802 - val_loss: 0.0909 - val_acc: 0.9737
Epoch 7/25
455/455 [==============================] - 0s 458us/step - loss: 0.0607 - acc:
0.9802 - val_loss: 0.1024 - val_acc: 0.9649
Epoch 8/25
455/455 [==============================] - 0s 455us/step - loss: 0.0585 - acc:
0.9846 - val_loss: 0.0999 - val_acc: 0.9737
Epoch 9/25
455/455 [==============================] - 0s 445us/step - loss: 0.0559 - acc:
0.9824 - val_loss: 0.0772 - val_acc: 0.9825
Epoch 10/25
455/455 [==============================] - 0s 461us/step - loss: 0.0527 - acc:
0.9868 - val_loss: 0.0842 - val_acc: 0.9737
Epoch 11/25
```

```
455/455 [==============================] - 0s 455us/step - loss: 0.0507 - acc:
0.9890 - val_loss: 0.0933 - val_acc: 0.9825
Epoch 12/25
455/455 [==============================] - 0s 454us/step - loss: 0.0500 - acc:
0.9846 - val_loss: 0.0961 - val_acc: 0.9825
Epoch 13/25
455/455 [==============================] - 0s 453us/step - loss: 0.0479 - acc:
0.9868 - val_loss: 0.1007 - val_acc: 0.9825
Epoch 14/25
455/455 [==============================] - 0s 451us/step - loss: 0.0453 - acc:
0.9868 - val_loss: 0.0948 - val_acc: 0.9737
Epoch 15/25
455/455 [==============================] - 0s 457us/step - loss: 0.0439 - acc:
0.9912 - val_loss: 0.0981 - val_acc: 0.9737
Epoch 16/25
455/455 [==============================] - 0s 453us/step - loss: 0.0417 - acc:
0.9912 - val_loss: 0.0901 - val_acc: 0.9737
Epoch 17/25
455/455 [==============================] - 0s 454us/step - loss: 0.0410 - acc:
0.9912 - val_loss: 0.0862 - val_acc: 0.9737
Epoch 18/25
455/455 [==============================] - 0s 455us/step - loss: 0.0395 - acc:
0.9912 - val_loss: 0.1044 - val_acc: 0.9649
Epoch 19/25
455/455 [==============================] - 0s 455us/step - loss: 0.0379 - acc:
0.9912 - val_loss: 0.1006 - val_acc: 0.9649
Epoch 20/25
455/455 [==============================] - 0s 458us/step - loss: 0.0371 - acc:
0.9912 - val_loss: 0.1051 - val_acc: 0.9649
Epoch 21/25
455/455 [==============================] - 0s 451us/step - loss: 0.0350 - acc:
0.9912 - val_loss: 0.0880 - val_acc: 0.9649
Epoch 22/25
455/455 [==============================] - 0s 454us/step - loss: 0.0348 - acc:
0.9890 - val_loss: 0.1087 - val_acc: 0.9649
Epoch 23/25
455/455 [==============================] - 0s 454us/step - loss: 0.0323 - acc:
0.9890 - val_loss: 0.1042 - val_acc: 0.9649
Epoch 24/25
455/455 [==============================] - 0s 457us/step - loss: 0.0322 - acc:
0.9890 - val_loss: 0.1121 - val_acc: 0.9561
Epoch 25/25
455/455 [==============================] - 0s 450us/step - loss: 0.0317 - acc:
0.9912 - val_loss: 0.1029 - val_acc: 0.9561
```

```
visualize_training_results(results_2)
```





What Happened?

Although the final validation score for both models is the same, this model is clearly worse because it hasn't converged yet. We can tell because of the greater variance in the movement of the `val_loss` and `val_acc` lines. This suggests that we can remedy this by either:

- Decreasing the size of the network, or
- Increasing the size of our training data

Visualizing why we Normalize our Data

As a final exercise, let's create a third model that is the same as the first model we created earlier. The only difference is that we will train it on our raw dataset, not the normalized version. This way, we can see how much of a difference normalizing our input data makes.

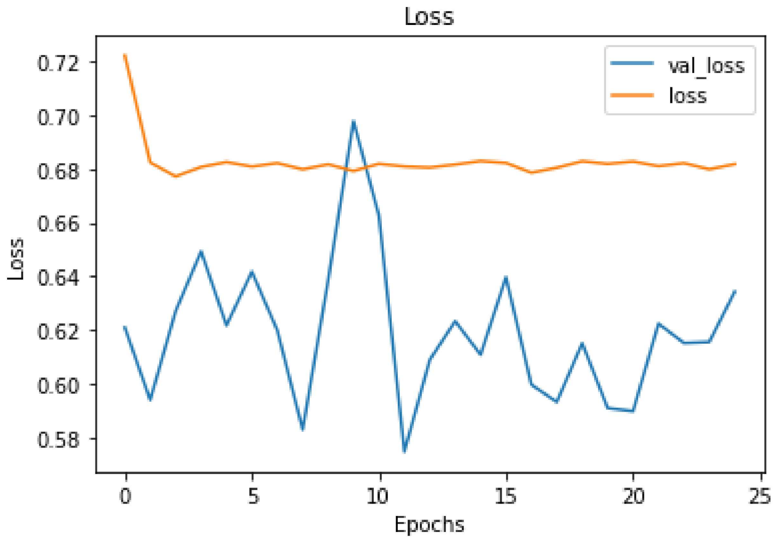
Create, compile, and fit a model in the cell below. The only change in parameters will be using `data` instead of `scaled_data` during the `.fit()` step.

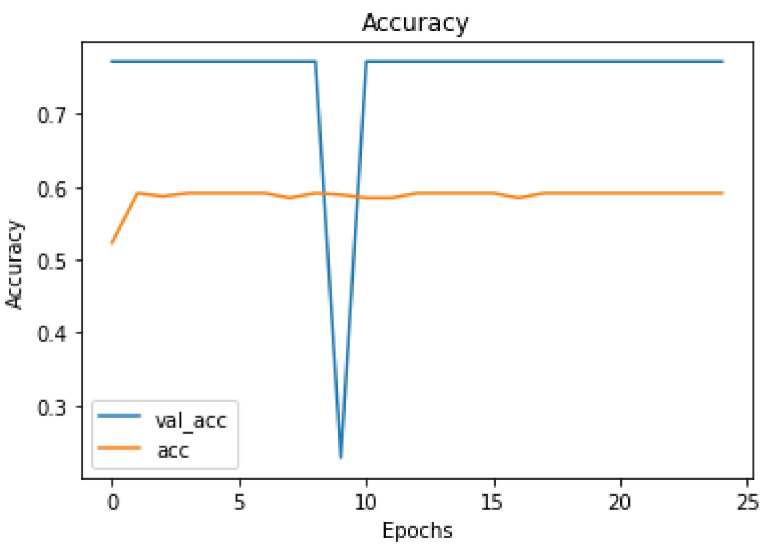
```
model_3 = Sequential()
model_3.add(Dense(5, activation='tanh', input_shape=(30,)))
model_3.add(Dense(1, activation='sigmoid'))



model_3.compile(loss='binary_crossentropy', optimizer='sgd', metrics=['acc'])



results_3 = model_3.fit(data, labels, epochs=25, batch_size=1, validation_split=0.2)
```

```
Epoch 1/25
455/455 [==============================] - 0s 634us/step - loss: 0.7220 - acc: 
0.5231 - val_loss: 0.6210 - val_acc: 0.7719
Epoch 2/25
455/455 [==============================] - 0s 444us/step - loss: 0.6823 - acc: 
0.5912 - val_loss: 0.5941 - val_acc: 0.7719
Epoch 3/25
455/455 [==============================] - 0s 439us/step - loss: 0.6771 - acc: 
0.5868 - val_loss: 0.6271 - val_acc: 0.7719
Epoch 4/25
455/455 [==============================] - 0s 446us/step - loss: 0.6806 - acc: 
0.5912 - val_loss: 0.6493 - val_acc: 0.7719
Epoch 5/25
455/455 [==============================] - 0s 428us/step - loss: 0.6825 - acc: 
0.5912 - val_loss: 0.6218 - val_acc: 0.7719
Epoch 6/25
455/455 [==============================] - 0s 437us/step - loss: 0.6808 - acc: 
0.5912 - val_loss: 0.6417 - val_acc: 0.7719
Epoch 7/25
455/455 [==============================] - 0s 447us/step - loss: 0.6821 - acc: 
0.5912 - val_loss: 0.6200 - val_acc: 0.7719
Epoch 8/25
455/455 [==============================] - 0s 438us/step - loss: 0.6798 - acc: 
0.5846 - val_loss: 0.5830 - val_acc: 0.7719
Epoch 9/25
455/455 [==============================] - 0s 426us/step - loss: 0.6816 - acc: 
0.5912 - val_loss: 0.6384 - val_acc: 0.7719
Epoch 10/25
455/455 [==============================] - 0s 438us/step - loss: 0.6792 - acc: 
0.5890 - val_loss: 0.6976 - val_acc: 0.2281
Epoch 11/25
455/455 [==============================] - 0s 439us/step - loss: 0.6818 - acc: 
```

```
0.5846 - val_loss: 0.6627 - val_acc: 0.7719
Epoch 12/25
455/455 [==============================] - 0s 435us/step - loss: 0.6808 - acc:
0.5846 - val_loss: 0.5749 - val_acc: 0.7719
Epoch 13/25
455/455 [==============================] - 0s 440us/step - loss: 0.6804 - acc:
0.5912 - val_loss: 0.6090 - val_acc: 0.7719
Epoch 14/25
455/455 [==============================] - 0s 435us/step - loss: 0.6815 - acc:
0.5912 - val_loss: 0.6233 - val_acc: 0.7719
Epoch 15/25
455/455 [==============================] - 0s 455us/step - loss: 0.6828 - acc:
0.5912 - val_loss: 0.6109 - val_acc: 0.7719
Epoch 16/25
455/455 [==============================] - 0s 442us/step - loss: 0.6821 - acc:
0.5912 - val_loss: 0.6397 - val_acc: 0.7719
Epoch 17/25
455/455 [==============================] - 0s 462us/step - loss: 0.6785 - acc:
0.5846 - val_loss: 0.5998 - val_acc: 0.7719
Epoch 18/25
455/455 [==============================] - 0s 443us/step - loss: 0.6804 - acc:
0.5912 - val_loss: 0.5933 - val_acc: 0.7719
Epoch 19/25
455/455 [==============================] - 0s 439us/step - loss: 0.6828 - acc:
0.5912 - val_loss: 0.6152 - val_acc: 0.7719
Epoch 20/25
455/455 [==============================] - 0s 438us/step - loss: 0.6818 - acc:
0.5912 - val_loss: 0.5910 - val_acc: 0.7719
Epoch 21/25
455/455 [==============================] - 0s 440us/step - loss: 0.6827 - acc:
0.5912 - val_loss: 0.5899 - val_acc: 0.7719
Epoch 22/25
455/455 [==============================] - 0s 435us/step - loss: 0.6810 - acc:
0.5912 - val_loss: 0.6224 - val_acc: 0.7719
Epoch 23/25
455/455 [==============================] - 0s 434us/step - loss: 0.6821 - acc:
0.5912 - val_loss: 0.6152 - val_acc: 0.7719
Epoch 24/25
455/455 [==============================] - 0s 440us/step - loss: 0.6798 - acc:
0.5912 - val_loss: 0.6156 - val_acc: 0.7719
Epoch 25/25
455/455 [==============================] - 0s 439us/step - loss: 0.6817 - acc:
0.5912 - val_loss: 0.6343 - val_acc: 0.7719
```

```
visualize_training_results(results_3)
```





Wow! Our results were much worse -- over 20% poorer performance when working with non-normalized input data!

Summary

In this lab, we got some practice creating ***Multi-Layer Perceptrons***, and explored how things like the number of layers in a model and data normalization affect our overall training results!

Releases

No releases published

Packages

No packages published

Contributors 6

Languages

- **Jupyter Notebook** 100.0%