lab_music_partial

November 22, 2022

1 Lab: Neural Networks for Music Classification

In addition to the concepts in the MNIST neural network demo, in this lab, you will learn to: * Load a file from a URL * Extract simple features from audio samples for machine learning tasks such as speech recognition and classification * Build a simple neural network for music classification using these features * Use a callback to store the loss and accuracy history in the training process * Optimize the learning rate of the neural network

To illustrate the basic concepts, we will look at a relatively simple music classification problem. Given a sample of music, we want to determine which instrument (e.g. trumpet, violin, piano) is playing. This dataset was generously supplied by Prof. Juan Bello at NYU Stenihardt and his former PhD student Eric Humphrey (now at Spotify). They have a complete website dedicated to deep learning methods in music informatics:

http://marl.smusic.nyu.edu/wordpress/projects/feature-learning-deep-architectures/deep-learning-python-tutorial/

You can also check out Juan's course.

1.1 Loading Tensorflow

Before starting this lab, you will need to install Tensorflow. If you are using Google colaboratory, Tensorflow is already installed. Run the following command to ensure Tensorflow is installed.

[ ]:
```
import tensorflow as tf
```

Then, load the other packages.

[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
```

1.2 Audio Feature Extraction with Librosa

The key to audio classification is to extract the correct features. In addition to `keras`, we will need the `librosa` package. The `librosa` package in python has a rich set of methods extracting the features of audio samples commonly used in machine learning tasks such as speech recognition and sound classification.

Installation instructions and complete documentation for the package are given on the librosa main page. On most systems, you should be able to simply use:

```
pip install librosa
```

For Unix, you may need to load some additional packages:

```
sudo apt-get install build-essential
sudo apt-get install libxext-dev python-qt4 qt4-dev-tools
pip install librosa
```

After you have installed the package, try to import it.

```
[ ]: import librosa
     import librosa.display
     import librosa.feature
```

```
c:\Users\user\anaconda3\lib\site-packages\paramiko\transport.py:219:
CryptographyDeprecationWarning: Blowfish has been deprecated
  "class": algorithms.Blowfish,
```

In this lab, we will use a set of music samples from the website:

http://theremin.music.uiowa.edu

This website has a great set of samples for audio processing. Look on the web for how to use the `requests.get` and `file.write` commands to load the file at the URL provided into your working directory.

You can play the audio sample by copying the file to your local machine and playing it on any media player. If you listen to it you will hear a soprano saxophone (with vibrato) playing four notes (C, C#, D, Eb).

```
[ ]: import requests
     fn = "SopSax.Vib.pp.C6Eb6.aiff"
     url = "http://theremin.music.uiowa.edu/sound files/MIS/Woodwinds/
      ↪sopranosaxophone/"+fn

     # TODO 1:  Load the file from url and save it in a file under the name fn
     response = requests.get(url)
     f = open(fn, 'wb')
     f.write(response.content)
     f.close()
```

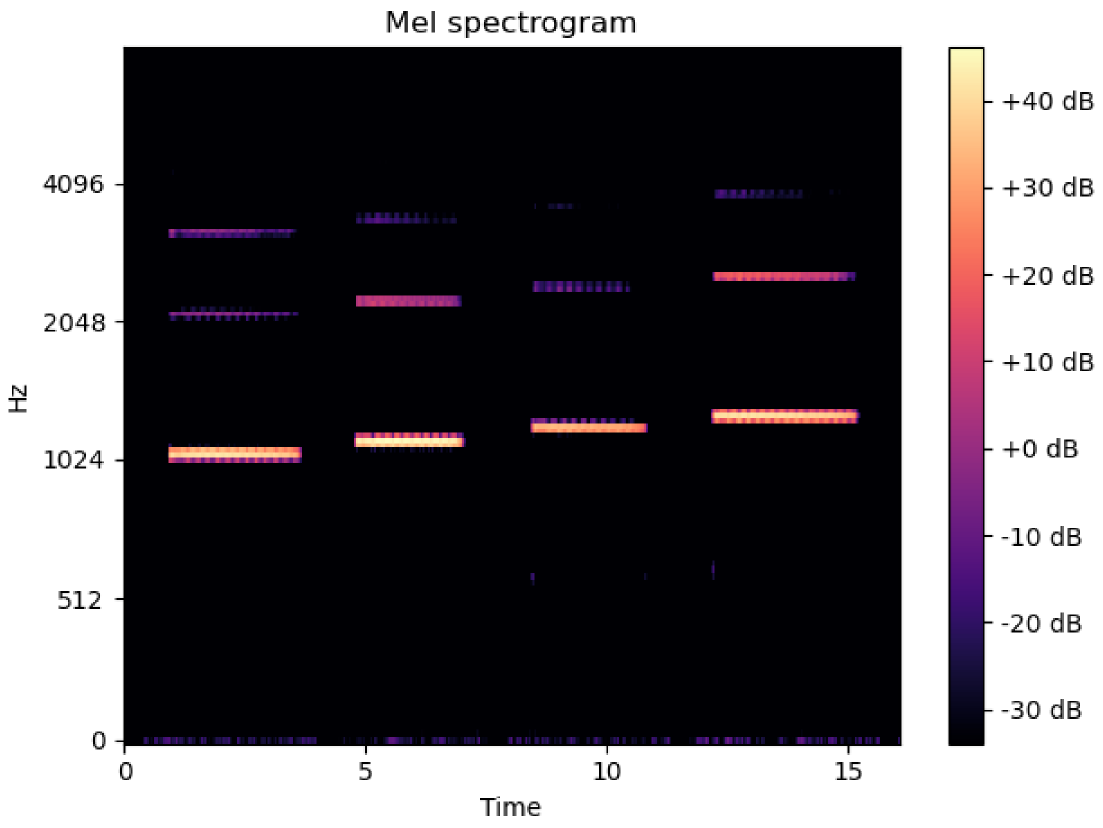
Next, use `librosa` command `librosa.load` to read the audio file with filename `fn` and get the samples `y` and sample rate `sr`.

```
[ ]: # TODO 2
     y, sr = librosa.load(fn)
```

Extracting features from audio files is an entire subject on its own right. A commonly used set of features are called the Mel Frequency Cepstral Coefficients (MFCCs). These are derived from the so-called mel spectrogram which is something like a regular spectrogram, but the power and frequency are represented in log scale, which more naturally aligns with human perceptual processing. You can run the code below to display the mel spectrogram from the audio sample.

You can easily see the four notes played in the audio track. You also see the 'harmonics' of each notes, which are other tones at integer multiples of the fundamental frequency of each note.

```
[ ]: S = librosa.feature.melspectrogram(y=y, sr=sr, n_mels=128, fmax=8000)
     librosa.display.specshow(librosa.amplitude_to_db(S),
                              y_axis='mel', fmax=8000, x_axis='time')
     plt.colorbar(format='%+2.0f dB')
     plt.title('Mel spectrogram')
     plt.tight_layout()
```



1.3 Downloading the Data

Using the MFCC features described above, Eric Humphrey and Juan Bellow have created a complete data set that can used for instrument classification. Essentially, they collected a number of data files from the website above. For each audio file, the segmented the track into notes and then extracted 120 MFCCs for each note. The goal is to recognize the instrument from the 120 MFCCs. The process of feature extraction is quite involved. So, we will just use their processed data provided at:

https://github.com/marl/dl4mir-tutorial/blob/master/README.md

Note the password. Load the four files into some directory, say `instrument_dataset`. Then, load them with the commands.

[ ]:
```
data_dir = 'instrument_dataset/'
Xtr = np.load(data_dir+'uiowa_train_data.npy')
ytr = np.load(data_dir+'uiowa_train_labels.npy')
Xts = np.load(data_dir+'uiowa_test_data.npy')
yts = np.load(data_dir+'uiowa_test_labels.npy')
```

Looking at the data files: * What are the number of training and test samples? * What is the number of features for each sample? * How many classes (i.e. instruments) are there per class?

[ ]:
```
# TODO 3
ntr, nf = Xtr.shape
nts = Xts.shape[0]
nclass = len(set(yts))
print('The number of training samples is {}'.format(ntr))
print('The number of test samples is {}'.format(nts))
print('the number of feaetures is {}'.format(nf))
print('There are {} classes'.format(nclass))
```

```
The number of training samples is 66247
The number of test samples is 14904
the number of feaetures is 120
There are 10 classes
```

Before continuing, you must scale the training and test data, `Xtr` and `Xts`. Compute the mean and std deviation of each feature in `Xtr` and create a new training data set, `Xtr_scale`, by subtracting the mean and dividing by the std deviation. Also compute a scaled test data set, `Xts_scale` using the mean and std deviation learned from the training data set.

[ ]:
```
# TODO 4: Scale the training and test matrices
mean = np.mean(Xtr, axis=0)
std = np.std(Xtr, axis=0)
Xtr_scale = (Xtr - mean) / std
Xts_scale = (Xts - mean) / std
```

1.4 Building a Neural Network Classifier

Following the example in MNIST neural network demo, clear the keras session. Then, create a neural network `model` with: * `nh=256` hidden units * `sigmoid` activation * select the input and output shapes correctly * print the model summary

[ ]:
```
from tensorflow.keras.models import Model, Sequential
from tensorflow.keras.layers import Dense, Activation
import tensorflow.keras.backend as K
```

[ ]:
```
# TODO 5 clear session
K.clear_session()
```

[ ]:
```
# TODO 6: construct the model
nh = 256
```

```
nout = 10
model = Sequential()
model.add(Dense(units=nh, input_shape=(nf,), activation='sigmoid',␣
 ↪name='hidden'))
model.add(Dense(units=nout, activation='softmax', name='output'))
```

[ ]:
```
# TODO 7:  Print the model summary
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 hidden (Dense)              (None, 256)               30976

 output (Dense)              (None, 10)                2570

=================================================================
Total params: 33,546
Trainable params: 33,546
Non-trainable params: 0
_________________________________________________________________
```

Create an optimizer and compile the model. Select the appropriate loss function and metrics. For the optimizer, use the Adam optimizer with a learning rate of 0.001

[ ]:
```
# TODO 8
from tensorflow.keras import optimizers
opt = optimizers.Adam(lr=0.001)
model.compile(optimizer=opt,
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])
```

```
c:\Users\user\anaconda3\lib\site-
packages\keras\optimizers\optimizer_v2\adam.py:114: UserWarning: The `lr`
argument is deprecated, use `learning_rate` instead.
  super().__init__(name, **kwargs)
```

Fit the model for 10 epochs using the scaled data for both the training and validation. Use the `validation_data` option to pass the test data. Also, pass the callback class create above. Use a batch size of 100. Your final accuracy should be >99%.

[ ]:
```
# TODO 9
hist = model.fit(Xtr_scale, ytr, epochs=10, batch_size=100,␣
 ↪validation_data=(Xts_scale,yts))
```

```
Epoch 1/10
663/663 [==============================] - 2s 3ms/step - loss: 0.3487 -
accuracy: 0.9064 - val_loss: 0.1736 - val_accuracy: 0.9607
Epoch 2/10
```

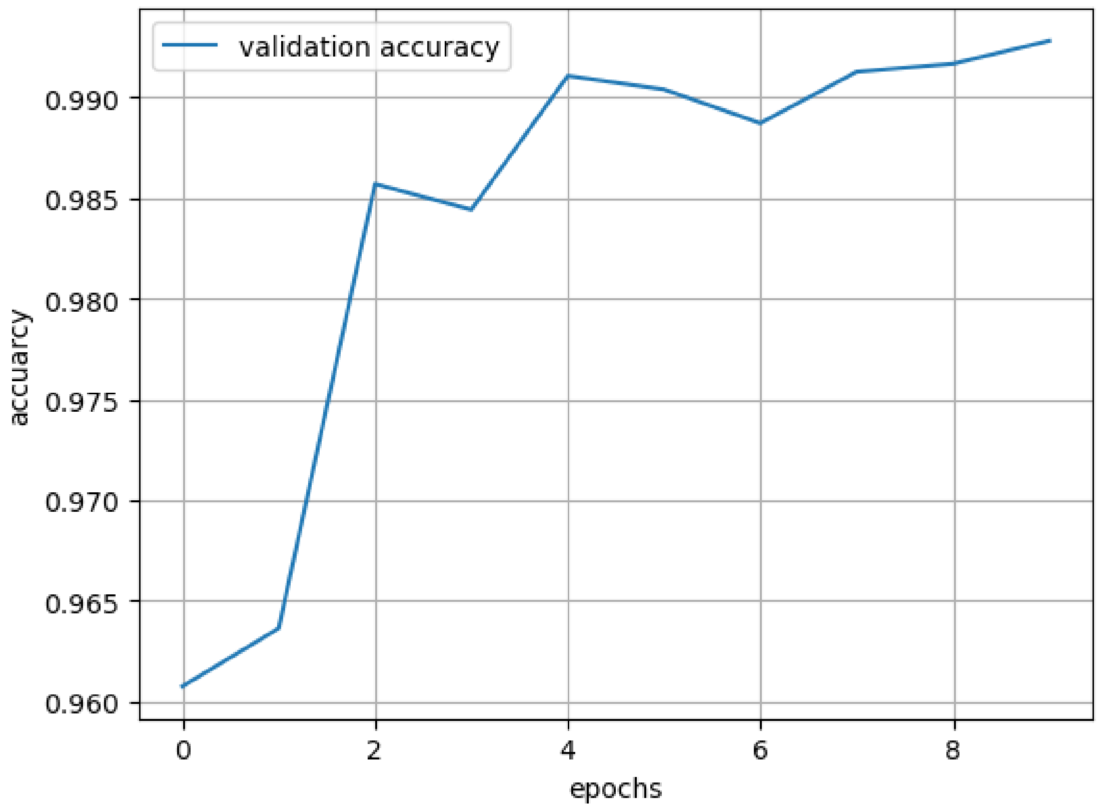
```
663/663 [==============================] - 2s 3ms/step - loss: 0.0991 -
accuracy: 0.9758 - val_loss: 0.1078 - val_accuracy: 0.9636
Epoch 3/10
663/663 [==============================] - 2s 3ms/step - loss: 0.0590 -
accuracy: 0.9857 - val_loss: 0.0584 - val_accuracy: 0.9857
Epoch 4/10
663/663 [==============================] - 2s 2ms/step - loss: 0.0415 -
accuracy: 0.9894 - val_loss: 0.0530 - val_accuracy: 0.9844
Epoch 5/10
663/663 [==============================] - 2s 2ms/step - loss: 0.0311 -
accuracy: 0.9920 - val_loss: 0.0375 - val_accuracy: 0.9911
Epoch 6/10
663/663 [==============================] - 2s 2ms/step - loss: 0.0249 -
accuracy: 0.9936 - val_loss: 0.0336 - val_accuracy: 0.9904
Epoch 7/10
663/663 [==============================] - 2s 3ms/step - loss: 0.0202 -
accuracy: 0.9949 - val_loss: 0.0375 - val_accuracy: 0.9887
Epoch 8/10
663/663 [==============================] - 2s 2ms/step - loss: 0.0172 -
accuracy: 0.9954 - val_loss: 0.0297 - val_accuracy: 0.9913
Epoch 9/10
663/663 [==============================] - 1s 2ms/step - loss: 0.0146 -
accuracy: 0.9961 - val_loss: 0.0275 - val_accuracy: 0.9917
Epoch 10/10
663/663 [==============================] - 2s 2ms/step - loss: 0.0124 -
accuracy: 0.9969 - val_loss: 0.0228 - val_accuracy: 0.9928
```

Plot the validation accuracy saved in `hist.history` dictionary. This gives one accuracy value per epoch. You should see that the validation accuracy saturates at a little higher than 99%. After that it "bounces around" due to the noise in the stochastic gradient descent.

[ ]:
```python
# TODO 10
val_accuracy = hist.history['val_accuracy']

plt.plot(val_accuracy)
plt.grid()
plt.xlabel('epochs')
plt.ylabel('accuarcy')
plt.legend(['validation accuracy'])
```

[ ]:
```
<matplotlib.legend.Legend at 0x1c2ad7c5460>
```

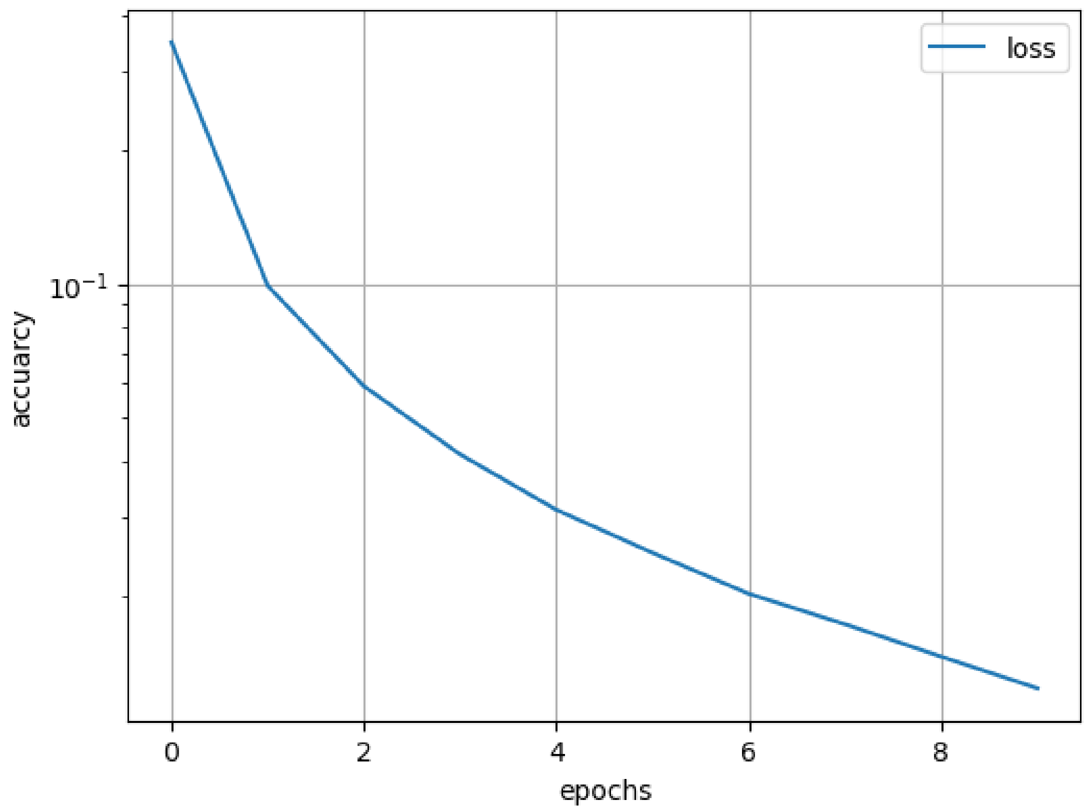


Plot the loss values saved in the `hist.history` dictionary. You should see that the loss is steadily decreasing. Use the `semilogy` plot.

```
[ ]: # TODO 11
     loss = hist.history['loss']

     plt.semilogy(loss)
     plt.grid()
     plt.xlabel('epochs')
     plt.ylabel('accuarcy')
     plt.legend(['loss'])
```

```
[ ]: <matplotlib.legend.Legend at 0x1c2ae0540d0>
```



1.5 Optimizing the Learning Rate

One challenge in training neural networks is the selection of the learning rate. Rerun the above code, trying four learning rates as shown in the vector `rates`. For each learning rate: * clear the session * construct the network * select the optimizer. Use the Adam optimizer with the appropriate learrning rate. * train the model for 20 epochs * save the accuracy and losses

[ ]:
```
rates = [0.01,0.001,0.0001]
batch_size = 100
acc_hist = []
loss_hist = []

# TODO 12
for lr in rates:
    K.clear_session()

    model = Sequential()
    model.add(Dense(units=nh, input_shape=(nf,), activation='sigmoid',␣
 ↪name='hidden'))
    model.add(Dense(units=nout, activation='softmax', name='output'))
    opt = optimizers.Adam(lr=lr)
    model.compile(optimizer=opt,
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

    hist = model.fit(Xtr_scale, ytr, epochs=20, batch_size=batch_size,␣
↪validation_data=(Xts_scale,yts))

    acc_hist.append(hist.history['val_accuracy'])
    loss_hist.append(hist.history['loss'])
```

```
Epoch 1/20
663/663 [==============================] - 2s 2ms/step - loss: 0.1070 -
accuracy: 0.9664 - val_loss: 0.0420 - val_accuracy: 0.9867
Epoch 2/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0296 -
accuracy: 0.9904 - val_loss: 0.0527 - val_accuracy: 0.9828
Epoch 3/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0234 -
accuracy: 0.9923 - val_loss: 0.0859 - val_accuracy: 0.9718
Epoch 4/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0167 -
accuracy: 0.9944 - val_loss: 0.0826 - val_accuracy: 0.9706
Epoch 5/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0154 -
accuracy: 0.9949 - val_loss: 0.0535 - val_accuracy: 0.9828
Epoch 6/20
663/663 [==============================] - 2s 3ms/step - loss: 0.0159 -
accuracy: 0.9951 - val_loss: 0.0862 - val_accuracy: 0.9803
Epoch 7/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0139 -
accuracy: 0.9957 - val_loss: 0.0469 - val_accuracy: 0.9848
Epoch 8/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0132 -
accuracy: 0.9957 - val_loss: 0.0925 - val_accuracy: 0.9754
Epoch 9/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0127 -
accuracy: 0.9960 - val_loss: 0.1062 - val_accuracy: 0.9740
Epoch 10/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0138 -
accuracy: 0.9957 - val_loss: 0.2494 - val_accuracy: 0.9554
Epoch 11/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0146 -
accuracy: 0.9957 - val_loss: 0.0597 - val_accuracy: 0.9822
Epoch 12/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0098 -
accuracy: 0.9969 - val_loss: 0.0551 - val_accuracy: 0.9840
Epoch 13/20
```

```
663/663 [==============================] - 1s 2ms/step - loss: 0.0067 -
accuracy: 0.9979 - val_loss: 0.0366 - val_accuracy: 0.9886
Epoch 14/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0111 -
accuracy: 0.9968 - val_loss: 0.0490 - val_accuracy: 0.9858
Epoch 15/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0096 -
accuracy: 0.9973 - val_loss: 0.0350 - val_accuracy: 0.9905
Epoch 16/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0100 -
accuracy: 0.9971 - val_loss: 0.0669 - val_accuracy: 0.9838
Epoch 17/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0086 -
accuracy: 0.9972 - val_loss: 0.0668 - val_accuracy: 0.9839
Epoch 18/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0068 -
accuracy: 0.9978 - val_loss: 0.0549 - val_accuracy: 0.9873
Epoch 19/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0084 -
accuracy: 0.9976 - val_loss: 0.1101 - val_accuracy: 0.9771
Epoch 20/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0104 -
accuracy: 0.9968 - val_loss: 0.0703 - val_accuracy: 0.9861
Epoch 1/20
663/663 [==============================] - 2s 2ms/step - loss: 0.3632 -
accuracy: 0.8993 - val_loss: 0.1915 - val_accuracy: 0.9471
Epoch 2/20
663/663 [==============================] - 1s 2ms/step - loss: 0.1028 -
accuracy: 0.9747 - val_loss: 0.1007 - val_accuracy: 0.9694
Epoch 3/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0602 -
accuracy: 0.9854 - val_loss: 0.0692 - val_accuracy: 0.9828
Epoch 4/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0425 -
accuracy: 0.9893 - val_loss: 0.0544 - val_accuracy: 0.9849
Epoch 5/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0319 -
accuracy: 0.9918 - val_loss: 0.0407 - val_accuracy: 0.9896
Epoch 6/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0253 -
accuracy: 0.9934 - val_loss: 0.0338 - val_accuracy: 0.9898
Epoch 7/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0210 -
accuracy: 0.9949 - val_loss: 0.0358 - val_accuracy: 0.9893
Epoch 8/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0173 -
accuracy: 0.9953 - val_loss: 0.0294 - val_accuracy: 0.9907
Epoch 9/20
```

```
663/663 [==============================] - 1s 2ms/step - loss: 0.0146 -
accuracy: 0.9962 - val_loss: 0.0258 - val_accuracy: 0.9924
Epoch 10/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0130 -
accuracy: 0.9966 - val_loss: 0.0257 - val_accuracy: 0.9924
Epoch 11/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0113 -
accuracy: 0.9973 - val_loss: 0.0263 - val_accuracy: 0.9909
Epoch 12/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0102 -
accuracy: 0.9973 - val_loss: 0.0268 - val_accuracy: 0.9901
Epoch 13/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0090 -
accuracy: 0.9978 - val_loss: 0.0262 - val_accuracy: 0.9905
Epoch 14/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0082 -
accuracy: 0.9977 - val_loss: 0.0228 - val_accuracy: 0.9911
Epoch 15/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0073 -
accuracy: 0.9983 - val_loss: 0.0204 - val_accuracy: 0.9925
Epoch 16/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0067 -
accuracy: 0.9983 - val_loss: 0.0204 - val_accuracy: 0.9926
Epoch 17/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0067 -
accuracy: 0.9982 - val_loss: 0.0264 - val_accuracy: 0.9913
Epoch 18/20
663/663 [==============================] - 2s 2ms/step - loss: 0.0060 -
accuracy: 0.9984 - val_loss: 0.0213 - val_accuracy: 0.9921
Epoch 19/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0052 -
accuracy: 0.9986 - val_loss: 0.0254 - val_accuracy: 0.9917
Epoch 20/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0050 -
accuracy: 0.9988 - val_loss: 0.0252 - val_accuracy: 0.9912
Epoch 1/20
663/663 [==============================] - 2s 2ms/step - loss: 1.0837 -
accuracy: 0.6648 - val_loss: 0.8465 - val_accuracy: 0.6846
Epoch 2/20
663/663 [==============================] - 1s 2ms/step - loss: 0.5405 -
accuracy: 0.8528 - val_loss: 0.5781 - val_accuracy: 0.8156
Epoch 3/20
663/663 [==============================] - 1s 2ms/step - loss: 0.3783 -
accuracy: 0.9100 - val_loss: 0.4337 - val_accuracy: 0.8760
Epoch 4/20
663/663 [==============================] - 1s 2ms/step - loss: 0.2946 -
accuracy: 0.9322 - val_loss: 0.3543 - val_accuracy: 0.8987
Epoch 5/20
```

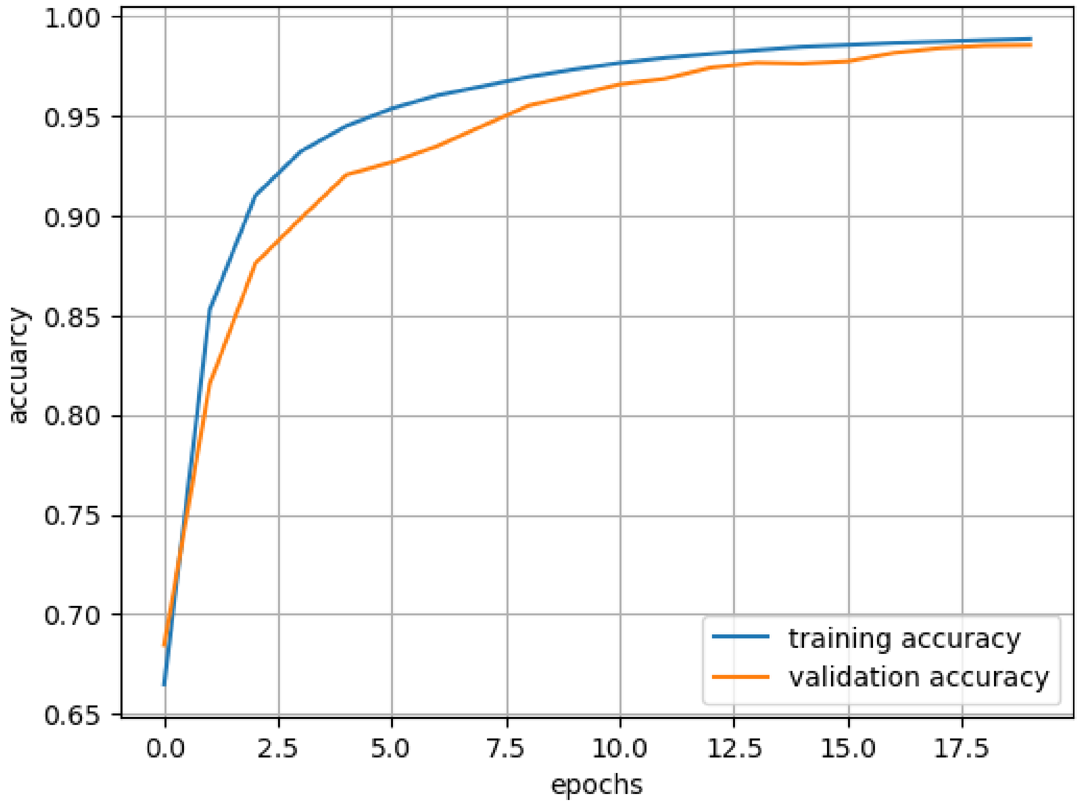
```
663/663 [==============================] - 1s 2ms/step - loss: 0.2408 -
accuracy: 0.9450 - val_loss: 0.2922 - val_accuracy: 0.9205
Epoch 6/20
663/663 [==============================] - 1s 2ms/step - loss: 0.2021 -
accuracy: 0.9537 - val_loss: 0.2518 - val_accuracy: 0.9268
Epoch 7/20
663/663 [==============================] - 2s 3ms/step - loss: 0.1724 -
accuracy: 0.9604 - val_loss: 0.2168 - val_accuracy: 0.9349
Epoch 8/20
663/663 [==============================] - 1s 2ms/step - loss: 0.1486 -
accuracy: 0.9649 - val_loss: 0.1854 - val_accuracy: 0.9451
Epoch 9/20
663/663 [==============================] - 1s 2ms/step - loss: 0.1293 -
accuracy: 0.9696 - val_loss: 0.1595 - val_accuracy: 0.9552
Epoch 10/20
663/663 [==============================] - 1s 2ms/step - loss: 0.1137 -
accuracy: 0.9733 - val_loss: 0.1393 - val_accuracy: 0.9604
Epoch 11/20
663/663 [==============================] - 1s 2ms/step - loss: 0.1007 -
accuracy: 0.9766 - val_loss: 0.1241 - val_accuracy: 0.9658
Epoch 12/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0901 -
accuracy: 0.9792 - val_loss: 0.1115 - val_accuracy: 0.9687
Epoch 13/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0812 -
accuracy: 0.9812 - val_loss: 0.0994 - val_accuracy: 0.9742
Epoch 14/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0739 -
accuracy: 0.9830 - val_loss: 0.0916 - val_accuracy: 0.9767
Epoch 15/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0676 -
accuracy: 0.9847 - val_loss: 0.0863 - val_accuracy: 0.9762
Epoch 16/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0623 -
accuracy: 0.9856 - val_loss: 0.0817 - val_accuracy: 0.9774
Epoch 17/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0578 -
accuracy: 0.9866 - val_loss: 0.0739 - val_accuracy: 0.9815
Epoch 18/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0538 -
accuracy: 0.9873 - val_loss: 0.0675 - val_accuracy: 0.9840
Epoch 19/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0503 -
accuracy: 0.9880 - val_loss: 0.0630 - val_accuracy: 0.9852
Epoch 20/20
663/663 [==============================] - 1s 2ms/step - loss: 0.0473 -
accuracy: 0.9887 - val_loss: 0.0616 - val_accuracy: 0.9855
```

Plot the loss funciton vs. the epoch number for all three learning rates on one graph. You should see that the lower learning rates are more stable, but converge slower.

```
[ ]: # TODO 13
tr_accuracy = hist.history['accuracy']
val_accuracy = hist.history['val_accuracy']

plt.plot(tr_accuracy)
plt.plot(val_accuracy)
plt.grid()
plt.xlabel('epochs')
plt.ylabel('accuarcy')
plt.legend(['training accuracy', 'validation accuracy'])
```

```
[ ]: <matplotlib.legend.Legend at 0x1c2be210910>
```



```
[ ]:
```

```
[ ]:
```