NEURAL NETS JEOPARDY!!

Bonus questions:

What is the name of John Connor's Father in Terminator?


What items does the terminator (Arnold) demand from the biker in the bar after he first 'lands' T2?
https://www.youtube.com/watch?v=r5qmFNoSYK4

What is the name of the upgrade Termintor in T2?
        https://www.youtube.com/watch?v=0uTO31B1R2c

What is the name of the robot in Short CIrcuit?
        https://www.youtube.com/watch?v=POxMp61Ksbk

Who is the lead character in Blade Runner?


Terminator
Blade Runner
Short Circuit
Ex Machina
West World
T2
RoboCop

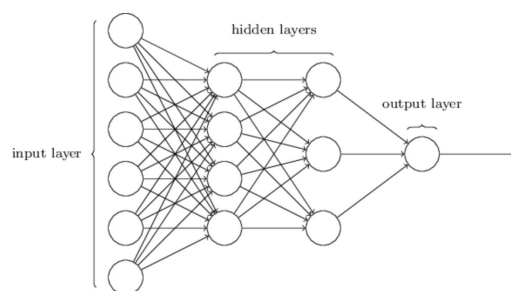----------------------------------------------------------------------------------------------------------------

100's

1.  Draw the basic architecture of a neural network:

2. How many output layers would we have for a regression problem? Classification?
   - One for regression
   - One for each class in a classification problem

3. What is an epoch?

4. As a general rule, it's good to start our first hidden layers with how many neurons?
       The same number as the number of features/columns we have in our data set

5. Neural Networks are notoriously high bias or high variance models?
       High variance


400:

1. How can we scale image data without using Standard Scaler?
   - By dividing by the maximum number of pixels in the image - this sets our pixels to between 0-1
2. What is the optimizer?
       "Adam"
3. When dealing with a multiclass classification problem, especially if our classes are numbers (like in the case of digit recognition), what do we need to do to our target variable?
   - Convert to dummies using to_categorical

4. Name the methods we discussed to prevent overfitting with NN:
Regularization (L1, L2)
Dropout
Early stopping

600:

1.  Name the 3 most common activation functions used in neural nets and explain when you'd use each one?
        Relu - hidden layers for either classification or regression
        Sigmoid- binary classification output
        Softmax - multiclass classification output
        None - regression output layer

2. What are the most commonly used loss functions in neural nets - name one for regression, binary class and multiclass classification:
        Regression - mean_squared_error
        Binary classification - binary cross entropy
        Multiclass classification - categorical cross entropy

3. After doing any essential data cleaning and EDA, what are the two things we need to do to our data before building our NN?
        TTS, Scale

4.  what dropout does:
It randomly drops out a specified proportion of nodes in each epoch. This is used to prevent overfitting.

5. Why is it a good idea to visualize your loss/accuracy and  val_loss/accuracy after you've run your model?

800s:

1.

Explain forward and back propagation:

Forward Propagation is straightforward -- either in batches or as individual observations, pass the training data through the network, applying all the weights, biases, and activation functions as usual. At this point, you should have actual and predicted values.

Back propagation:

1. See how far off we were from the truth using the loss function
2. Identify which weights in our network are most responsible for how far we are off
3. Change all of the weights to make our model more accurate, changing the weights that are "the worst" the most

This is known as **Backpropagation** -- we are taking the errors we see in our model (as it stands currently) and are distributing them backwards to the rest of the layers.

- **looks at how badly we did on each pass, moves those errors back up the model, and then uses gradient descent to change the weight over a series of iterations**.

2. Explain what early stopping does and what it's used for:

Used to prevent overfitting. Instead of continuing training through every epoch, once the validation error begins to increase, our algorithm stops because it has (in theory) found the minimum for the validation loss.

3. Convolutional NNs consist of 3 types of layers, what are they?

- Convolutional Layers
- Pooling layers
- Densely connected layers

4. In basic terms, what does a CNN do?

Takes an image, compresses it in a specific way, then connects it to densely connected layers

- They can be used t detect 'things' in objects
- Early layers detect edges (edge of a nose)
- Later layers detect structures (a nose)
- Final layers detect entire objects (faces)

1000s

1. In a CNN, as your filter size increases, what happens to the bias variance trade off in your model?

- As your filter size increases, you compress the image more, your result will be smaller, so there are less parameters to learn. If there are fewer parameters, there is less variance and an increase in bias (because we're reducing complexity)

2. What is padding?

When you add a layer of zeros around your image data - so you don't lose the data from the corners of your image

3. What is stride?

How steps to move the filter (default is 1)

4. Calculate the output of the image with this pixel matrix:

- There is no padding, this filter is run over it with a stride of 1 [-1,1, 0

[-1,1 0

-1.1, 0]

And max pooling is applied, with pooling size 2x2

[[ 4 5 5 6

3 5 6 7

3 4 5 6 ]]

Final Jeopardy:

      On your laptop (in a jupyter notebook), write out the code you would use to build a neural net for a regression problem in keras that. The size of the X_train data is (56000, 9).  Be sure to include the model architecture, the compiler and fitting the model.

Has 2 hidden layers
Includes 0.5 dropout
One output layer
Uses 10 epochs
Batch size = 50

```
model = Sequential()


model.add(Dense(9, input_dim=X_train.shape[1, activation='relu'))
model.Dropout(0.5)
model.add(Dense(5, activation= 'relu)
model.add(Dense(1, activation=None)

model.compile(loss='mean_squared_error', optimizer='adam')

model.fit(X_trian, y_train, validation_data = (X_test, y_test), epochs = 10)
```