1. I believe It's the mean and standard deviation for each channel. So that means that the average value should be 0.5, with an average distance from that being + or - 0.5. So that will help normalize each channel to a value 0 to 1.

2. Each Conv2d takes in the number of inputs, number of outputs, and the kernel size.

3. The MaxPool2d simplifies the layers by reducing their dimensionality/size. This makes computation easier. It takes the kernel size and stride, which determines the exact output and resulting size.

4. This actually runs the features and layers through the Neural Network. So I believe this is the part that actually takes the inputs and multiplies them by the weights (and adds the bias). And the inputs are just the amount of input and output features .

5. This removes all the dimensions. This is the way that the program needs the inputs to be able to read them through the network.

6. So I didn't create a separate script, but instead I check for a saved network, and then load it if it exists automatically. Otherwise it will train the network from the datasets.