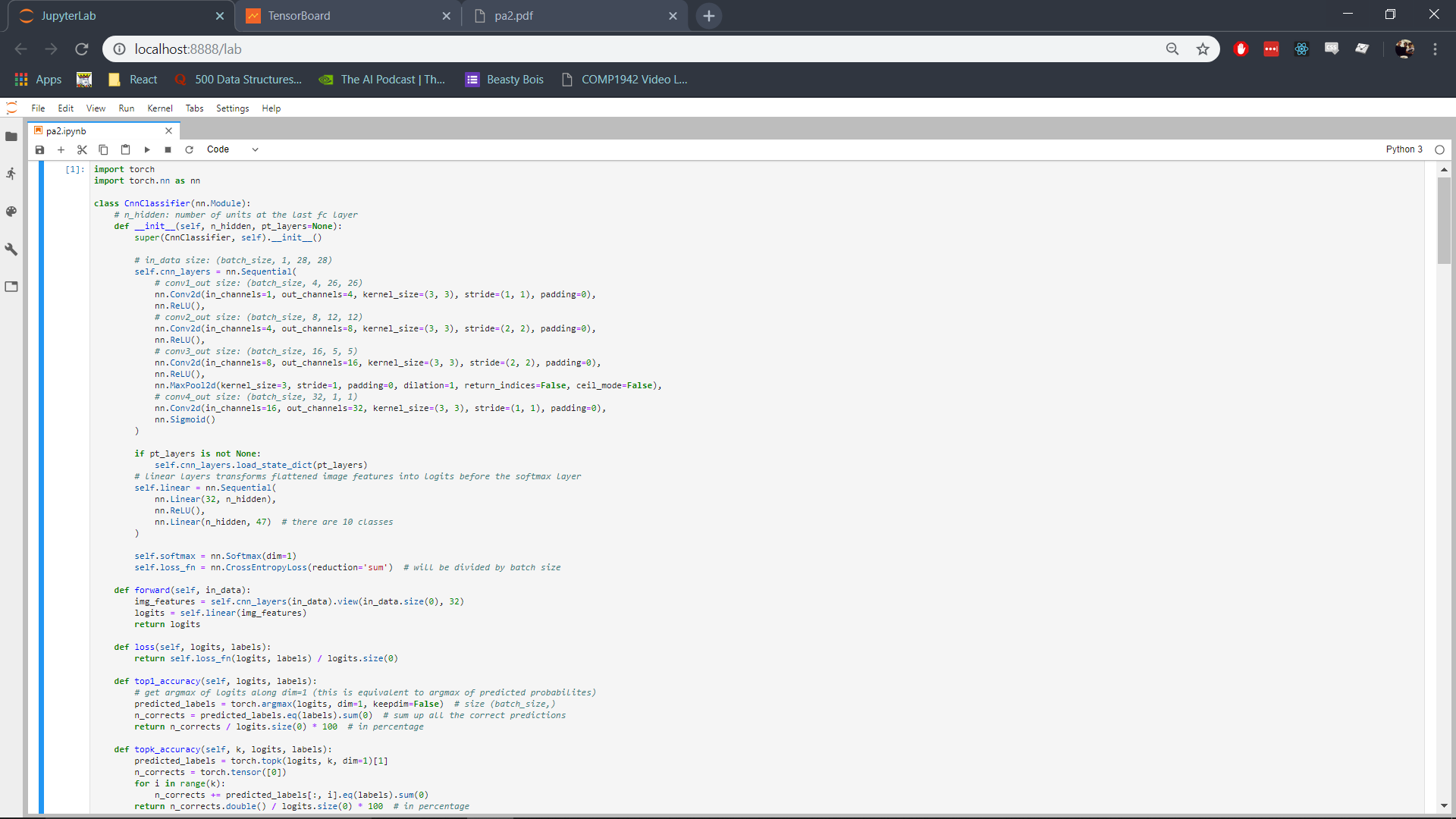
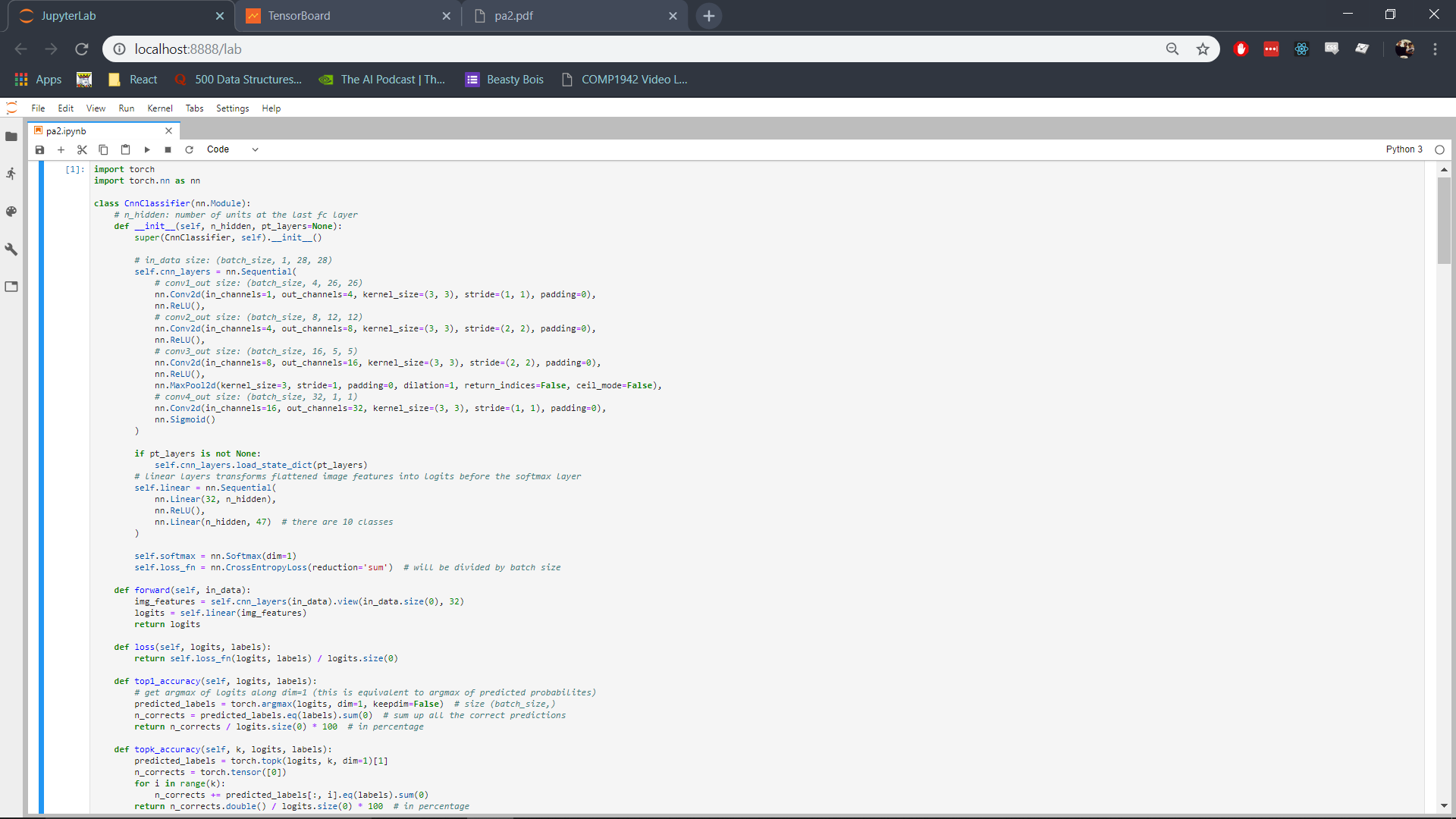
# CNN Classifiers

## Class for CNN Classifier

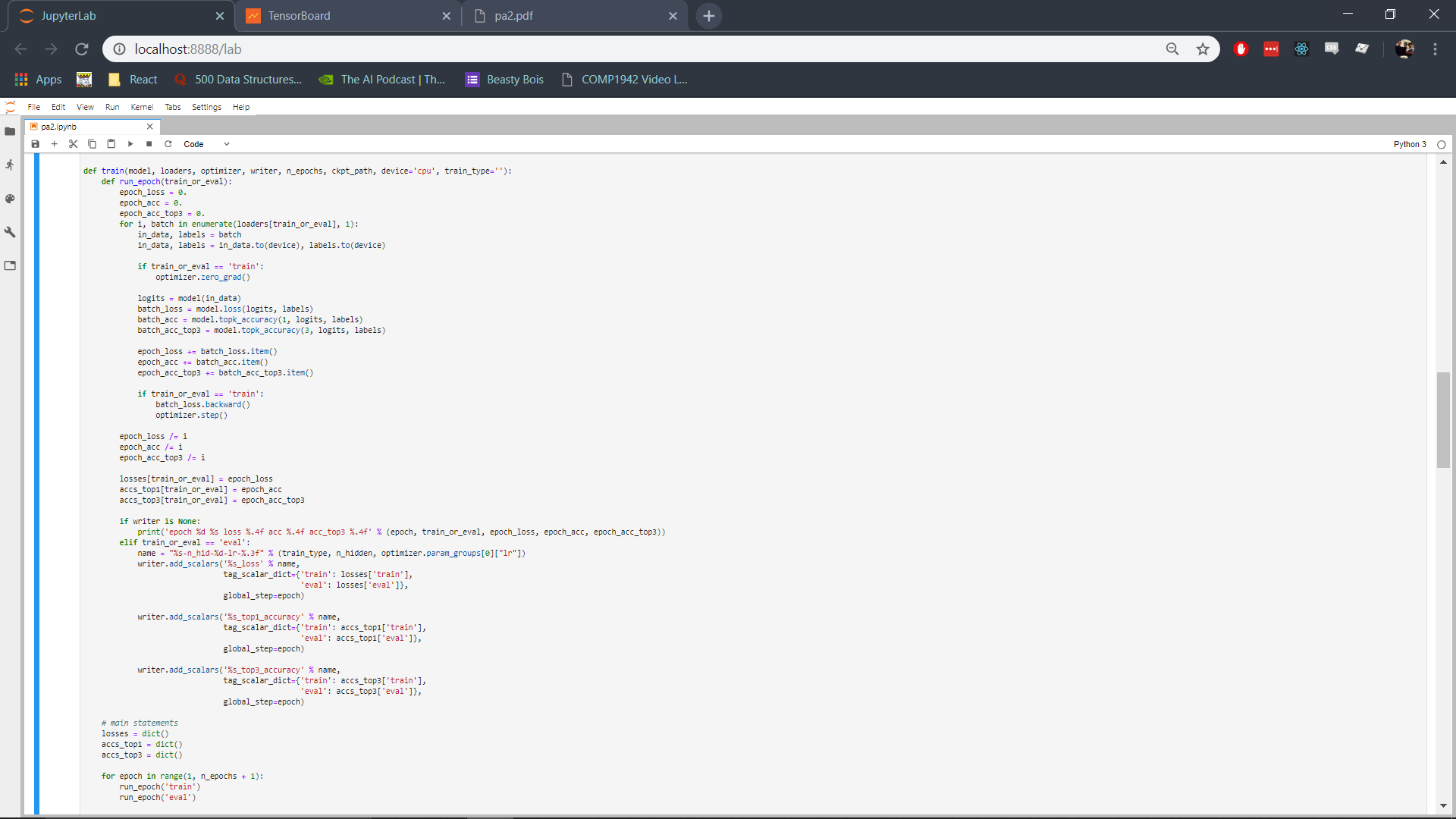


This class is used for making a classifier from scratch and to initialize the weights in case a pretrained network’s state\_dict is supplied in the constructor as the pt\_layers argument. The weights are loaded using the load\_state\_dict function of the nn.Sequential class as show below.

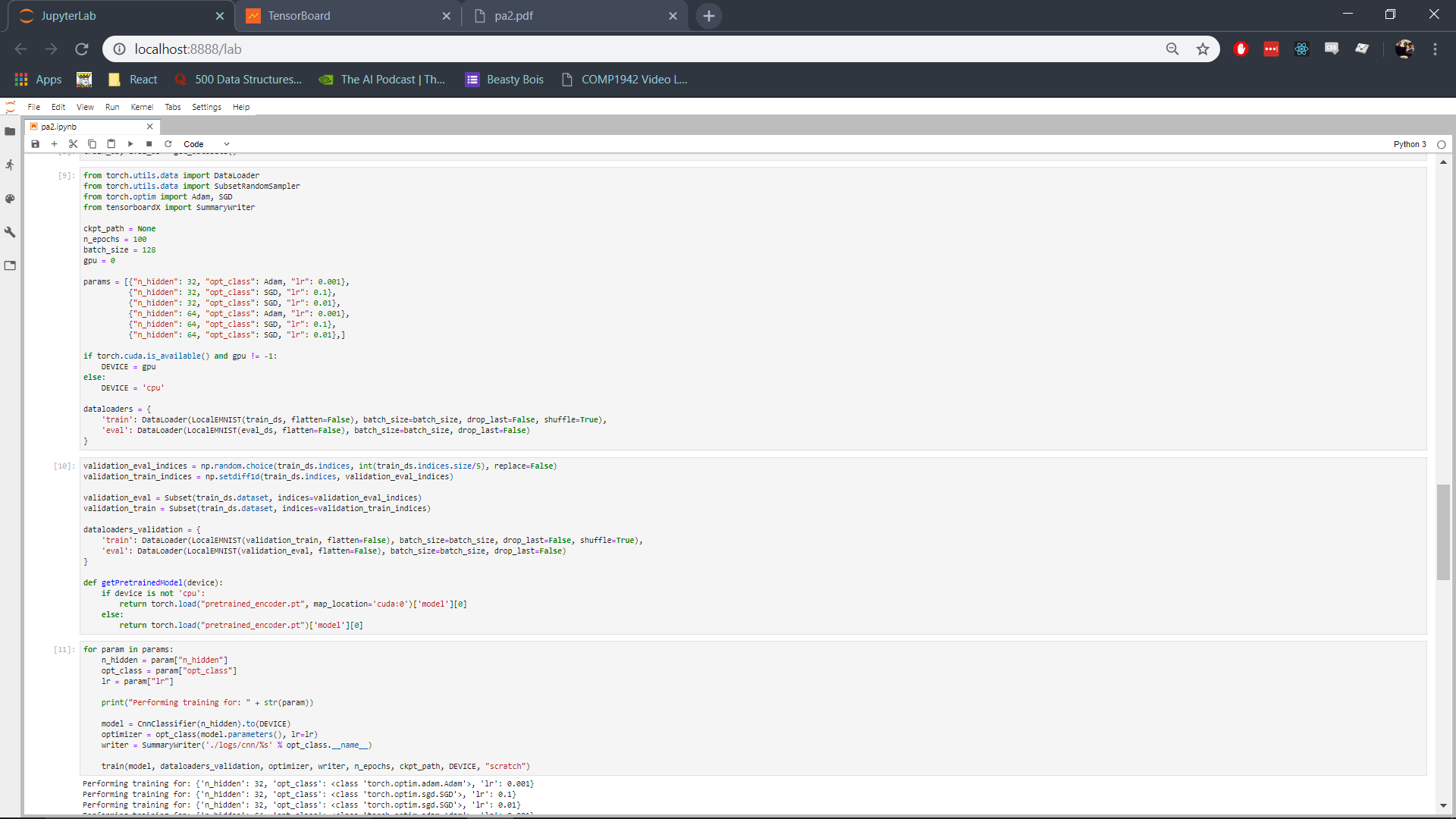


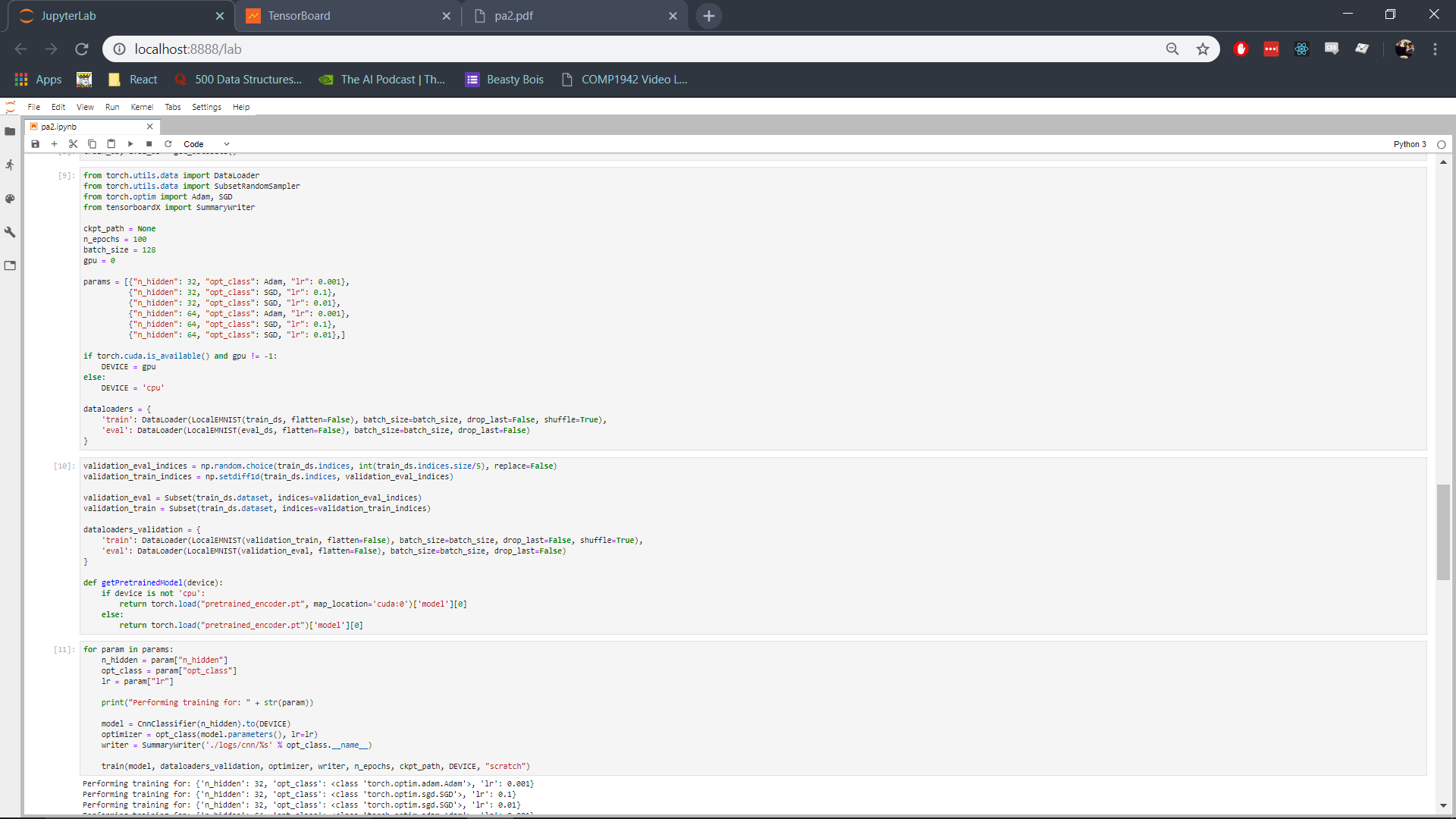
More details for loading the state\_dict from the pretrained layers are given below.

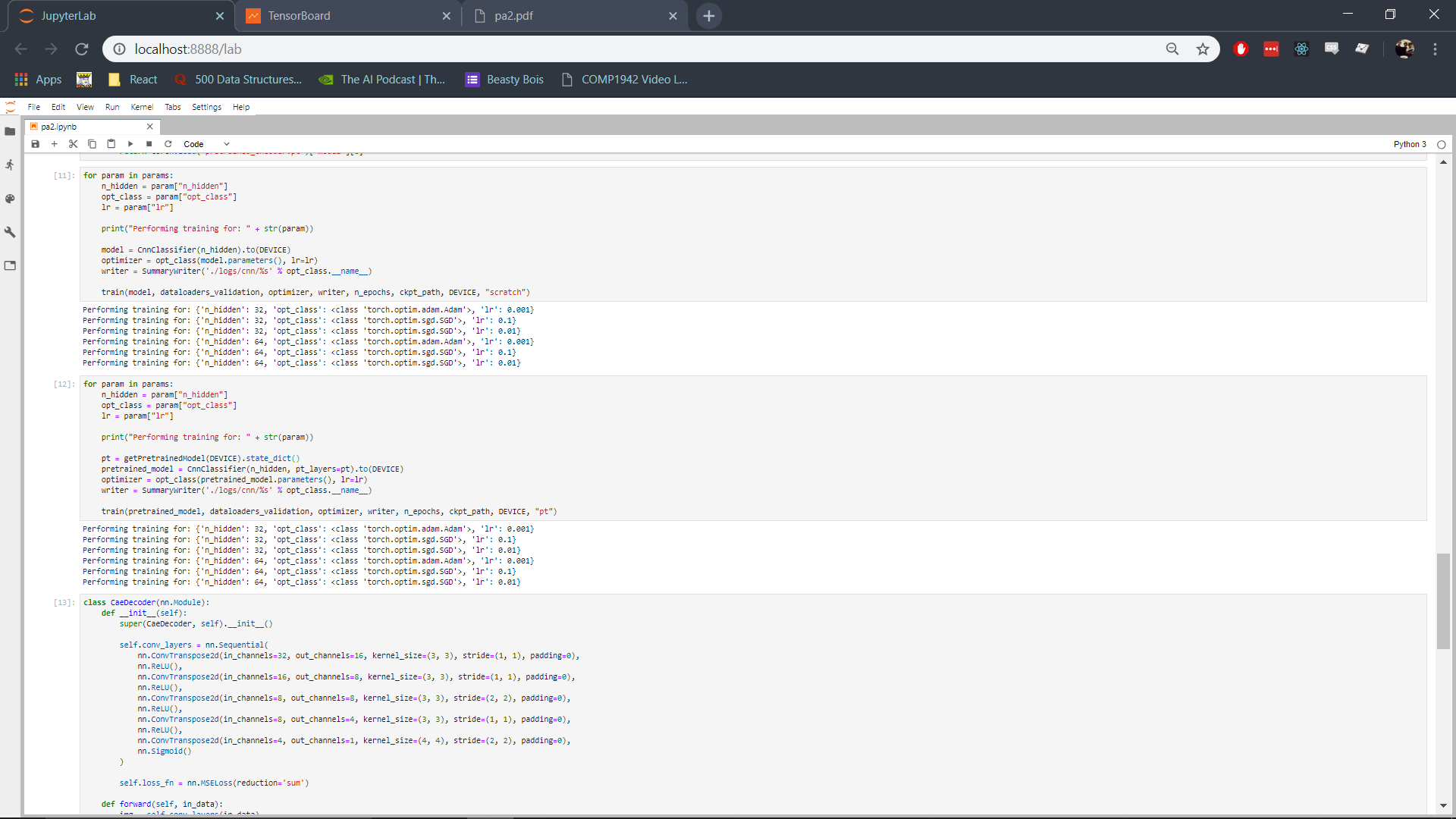
## CNN from Scratch and PT Train

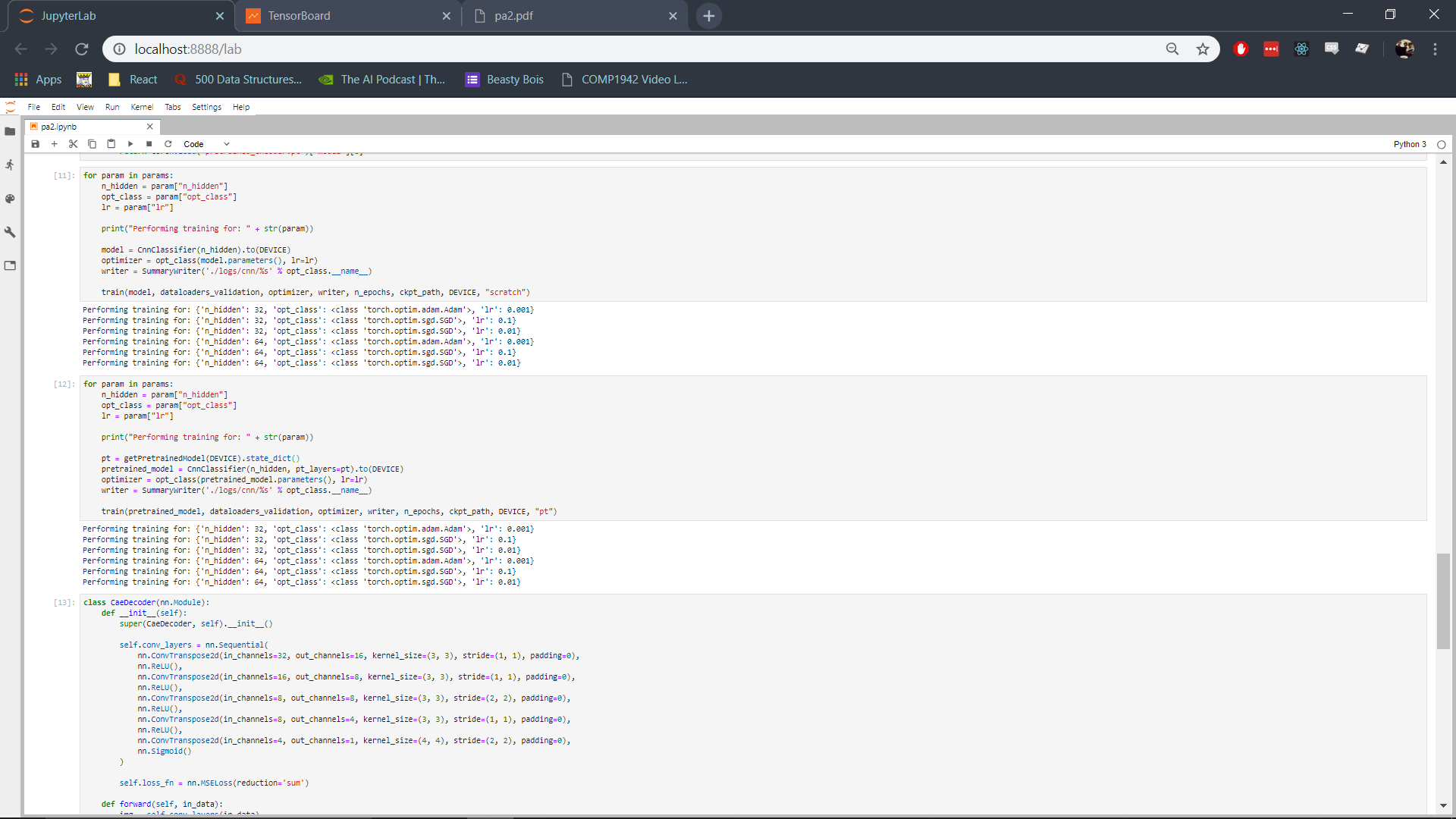


Run CNNs









## Validation Results

### Training from Scratch

|  |  |  |  |
| --- | --- | --- | --- |
| ***H*** | 32 | 32 | 32 |
| **Optimizer** | Adam | SGD | SGD |
| **Learning Rate** | 0.001 | 0.01 | 0.1 |
| **Loss** |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| ***H*** | 64 | 64 | 64 |
| **Optimizer** | Adam | SGD | SGD |
| **Learning Rate** | 0.001 | 0.01 | 0.1 |
| **Loss** |  |  |  |

### Training with Pretrained weights

|  |  |  |  |
| --- | --- | --- | --- |
| ***H*** | 32 | 32 | 32 |
| **Optimizer** | Adam | SGD | SGD |
| **Learning Rate** | 0.001 | 0.01 | 0.1 |
| **Loss** |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| ***H*** | 64 | 64 | 64 |
| **Optimizer** | Adam | SGD | SGD |
| **Learning Rate** | 0.001 | 0.01 | 0.1 |
| **Loss** |  |  |  |

## Details for training from scratch

From the results of the validation stage the best hyperparameters must be chosen. The SGD optimizer with a learning rate of 0.01 can be easily discarded as, from the loss curves for both *H* values, it can be seen that this setting does not perform well. When 32 is set for the value of *H* the network is unable to meaningfully train, with loss staying at its initial value of 3.85 even after 100 epochs of training. With 64 hidden layers the model performs better but the training is still unpredictable; it stays at a loss value of 3.86 for around 80 epochs and then it finally decreases to 1.5 at the hundredth epoch. It is possible that the loss would have decreased even more if trained for more than 100 epochs but it taking this long to even start decreasing in loss while other parameters perform much better already makes using it infeasible. The best validation loss value for 32 *H* is 3.85 and for 64 *H* it is 1.537.

SGD works much faster with a learning rate of 0.1, the loss remains steady at 3.85 for 20 epochs when using 32 for *H* and for only 10 epochs when using 64 for *H* after which it decreases in both cases to a final value of around 0.5. While this is much better than the previously mentioned setting, it is outperformed by the final hyperparameter setting which used Adam. The best validation loss value for 32 *H* is 0.5471 and for 64 *H* it is 0.5532.

The Adam optimizer with a learning rate of 0.001 clearly outperforms the other settings with it resulting in an immediate decrease in the value for loss, reaching a value under 1 in just 5 epochs, and becoming steady at around 0.5 in 40 epochs. Both values for *H* perform very similarly so choosing either would result in comparable performance, however 64 performs marginally better with its best loss value 0.5474 while the best loss value for 32 *H* is 0.6061

CAE with Pretrained Encoder 