

Classify the Adult dataset

I. Network structure

The network structure is as follows:

Output layer: Fully connected layer 2 with 2 output units.
Fully connected layer 1 with 128 output units and ReLU activation followed by BatchNorm1d and Dropout.
Input layer

The loss function is crossentropy because this network is doing classification tasks. Only fully connected layer is used in this problem. The dropout is utilized to improve the generality of the model. Given the loss function (crossentropy), the softmax activation function is implicitly chosen because this represents the probability of belonging to each class well.

II. Representative weights

FC1:

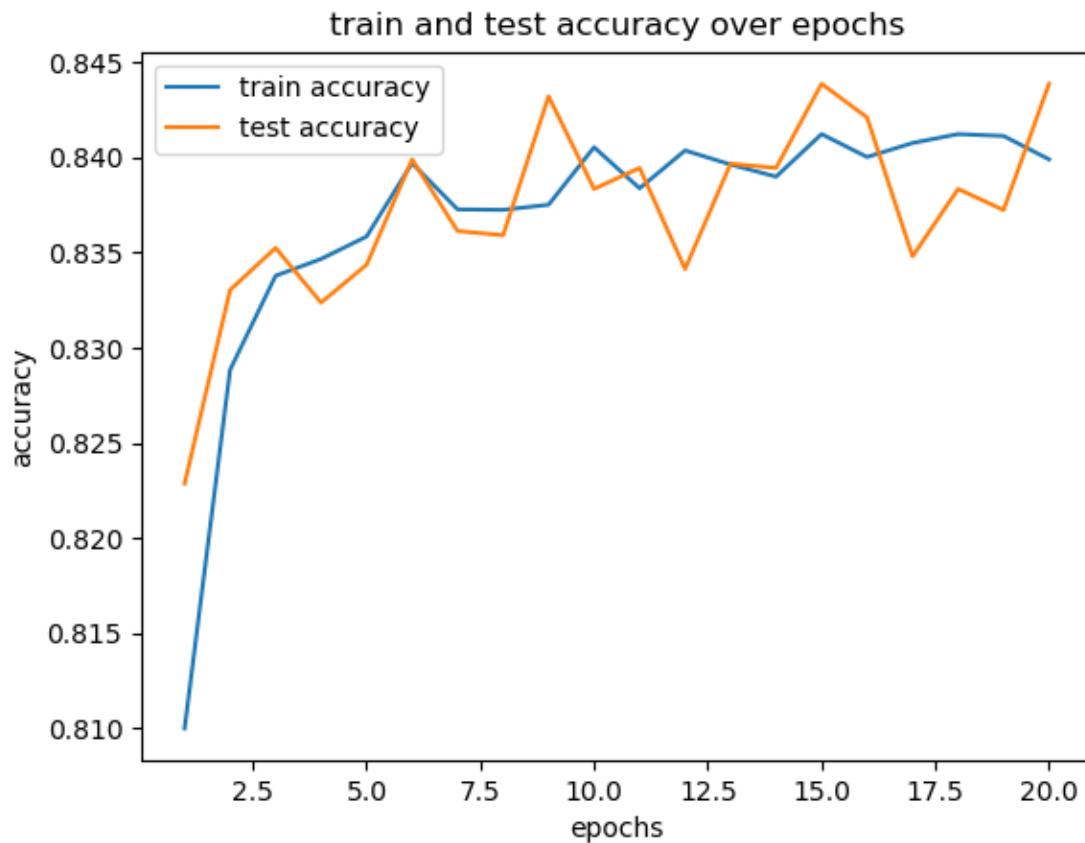
```
tensor([ 6.5721e-43,  6.5441e-43, -4.0778e-43,  6.6702e-43, -1.3450e-34,  
        -1.4045e-37, -6.5861e-43, -1.6956e-43,  5.1848e-44, -6.6702e-43,  
        6.3479e-43,  6.8523e-43,  6.6562e-43, -6.6842e-43,  4.4281e-43,  
        1.9058e-43, -5.7593e-43, -6.7963e-43, -6.6702e-43, -6.5020e-43],  
        device='cuda:0', grad_fn=<SliceBackward>)
```

FC2:

```
tensor([-0.0709, -0.2196, -0.2387, -0.0862,  0.0123, -0.0158,  0.0333, -0.0054,  
        -0.0717, -0.0843, -0.1515,  0.0480,  0.1696, -0.0681,  0.0974,  0.0157,  
        -0.0112, -0.2203, -0.1683, -0.0329],  
        device='cuda:0', grad_fn=<SliceBackward>)
```

As we can see from the above prints, all of the weights are in the range $[-1, 1]$, which is a reasonable range. The weights are not tuned to be all zeros or identical, so they are likely to yield a good output.

III. Training and testing accuracy vs epoch



The training accuracy is increasing with the number of epochs, which is expected. Because of the dropouts, the model generalizes well so the test accuracy is also increasing at first few epochs. Then the test accuracy is oscillating around 84%. Because of the simplicity of the network and the dropout layers, the train accuracy is also around 84% at the end.