

Classify the Breast Cancer dataset

I. Network structure

The network structure is as follows:

Output layer: Fully connected layer with 2 units and Softmax activation
Dropout layer with dropout rate 0.5
Fully connected layer with 128 units and ReLU activation
Dropout layer with dropout rate 0.25
Fully connected layer with 32 units and ReLU activation
Input layer

The loss function is crossentropy because this network is doing classification tasks. The fully connected layer is used for the first layer because the input is simply an array of 9 numbers. The dropout layer is there to prevent the model from overfitting. Then, there is a fc layer with 128 units with ReLU activation. Another dropout layer after the first fc layer is added to enhance the performance on test data. Finally, a fc layer with 2 units is the output layer. The softmax activation function is chosen there because this represents the probability of belonging to each class well.

II. Representative weights

FC1:

[-0.07823225, -0.00316991, 0.18374959, -0.25072658, -0.16153084,
-0.33726153, -0.03876471, 0.11905467, 0.13607539, 0.18258055,
-0.35013688, -0.1114786, -0.15397695, -0.12903094, 0.08494836,
-0.28544906, -0.02617798, 0.1442003, 0.0327351, -0.20973077,
-0.04867353, 0.04646942, 0.03687911, -0.09992623, -0.0212763,

0.21294409, 0.16583572, 0.06469265, -0.230964 , -0.09536251,
0.0435145 , -0.36298546]

FC2:

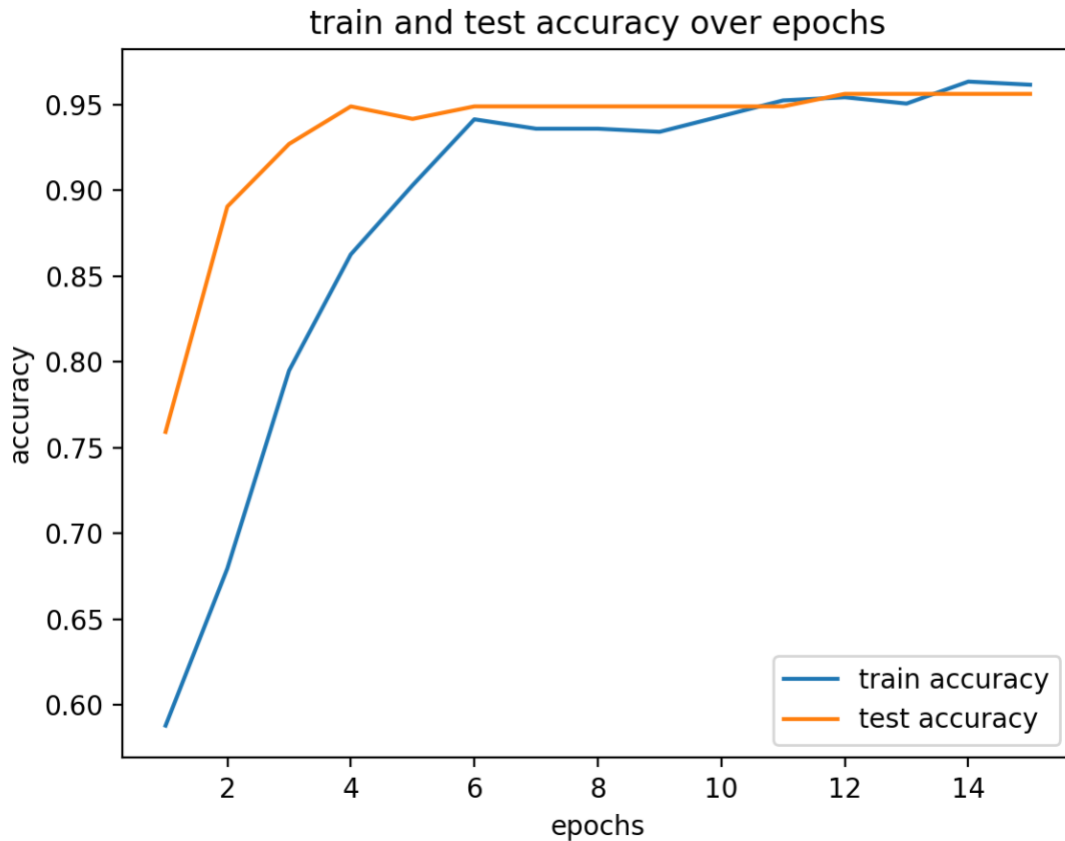
[-0.00255071, 0.14548156, 0.02310037, ..., -0.07189824,
0.21683797, 0.07874739],
[-0.13553952, -0.01447164, -0.08288918, ..., -0.07999652,
-0.13232556, 0.05835263],
[0.12962714, -0.17256977, 0.18505967, ..., 0.05362 ,
0.13164131, 0.13187526],
...,
[-0.10260877, 0.05500673, 0.15707387, ..., -0.0173333 ,
0.00026207, -0.0510395]

FC3:

[-1.05086658e-02, 1.04758017e-01],
[1.19034566e-01, 8.24897066e-02],
[-2.19588891e-01, -2.95663569e-02],
[-1.29032686e-01, 5.06069213e-02],
[-2.27269277e-01, 1.23796202e-02],
[2.38512486e-01, -1.14833549e-01],
[2.97518969e-01, -2.40840375e-01],
[2.28284998e-03, 2.40040980e-02],
[2.10092351e-01, -3.43921989e-01],
[-2.26314038e-01, -9.81204659e-02],
[-1.14150986e-01, 8.48423466e-02],
[-6.04974441e-02, 1.03224963e-01],
[-9.88328606e-02, 2.11496398e-01],
[7.83639103e-02, -1.21801607e-01],
[1.39158279e-01, 9.30843130e-03],
[1.61267430e-01, 6.51376368e-03]

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 higher than the training accuracy at first few epochs. However, the test accuracy cannot be improved in later epochs because of the overfitting and the final test accuracy is about 96%.

IV. Training and test data separation

The training data is selected to be first 4/5 of the breast cancer data, and the remaining 1/5 is chosen to be the test data. Because no prior information is known to the given data set, this division can be regarded as random.

V. Data points that fail in the model

```
[0.4 0.3 0.1 0.1 0.2 0.1 0.4 0.8 0.1]
[0.5 0.4 0.5 0.1 0.8 0.1 0.3 0.6 0.1]
[0.4 0.1 0.1 0.3 0.1 0.5 0.2 0.1 0.1]
[0.5 0.7 0.7 0.1 0.5 0.8 0.3 0.4 0.1]
[0.5 0.4 0.4 0.5 0.7 1. 0.3 0.2 0.1]
```

Note the data points have been scaled to range $[0,1]$. Because this data input is not a picture, we cannot visualize it. We don't know any prior information about the meaning of the data, so we are not able to explain why these data points failed. On the other hand, we can achieve 96% accuracy in test data set, which is a very good result. This indeed demonstrates the advantage of deep learning.