

DEEP LEARNING HOMEWORK 1

ALPTEKIN ORBAY 201700090

CONTENTS

1. Forward	2
1.1. Open Form	2
1.2. Compact Form	2
2. Backward Derivation	3
2.1. Open Form	3
2.2. Compact Form	4
3. Experiments	4
4. Source Code	6
5. Additional Material	11

1. FORWARD

1.1. Open Form.

$$E = -r \log(z) - (1 - r) \log(1 - z)$$

$$z = \sigma(\bar{z})$$

$$\bar{z} = u_1 h_1 + u_2 h_2 + b_z$$

$$h_1 = \sigma(\bar{h}_1)$$

$$h_2 = \sigma(\bar{h}_2)$$

$$\bar{h}_1 = w_{11}x_1 + w_{12}x_2 + b_1$$

$$\bar{h}_2 = w_{21}x_1 + w_{22}x_2 + b_2$$

1.2. Compact Form.

$$E = -r \log(z) - (1 - r) \log(1 - z)$$

$$z = \sigma(\bar{z})$$

$$\bar{z} = \mathbf{U}\mathbf{h} + b_z$$

$$\mathbf{h} = \sigma(\bar{\mathbf{h}})$$

$$\bar{\mathbf{h}} = \mathbf{W}\mathbf{x} + \mathbf{b}_h$$

2. BACKWARD DERIVATION

2.1. Open Form.

$$\begin{aligned}
\frac{\partial E}{\partial z} &= \frac{z-r}{z(1-z)} & \frac{\partial z}{\partial \bar{z}} &= z(1-z) & \frac{\partial E}{\partial \bar{z}} &= z-r \\
\frac{\bar{z}}{\partial u_1} &= h_1 & \frac{\bar{z}}{\partial u_2} &= h_2 & \frac{\bar{z}}{\partial b_z} &= 1 \\
\frac{\bar{z}}{\partial h_1} &= u_1 & \frac{\bar{z}}{\partial h_2} &= u_2 \\
\frac{h_1}{\partial h_1} &= h_1(1-h_1) & \frac{h_2}{\partial h_2} &= h_2(1-h_2) \\
\frac{\partial \bar{h}_1}{w_{11}} &= x_1 & \frac{\partial \bar{h}_1}{w_{12}} &= x_2 & \frac{\partial \bar{h}_2}{w_{21}} &= x_1 & \frac{\partial \bar{h}_2}{w_{22}} &= x_2 \\
\frac{\partial \bar{h}_2}{b_1} &= 1 & \frac{\partial \bar{h}_2}{b_2} &= 1 \\
\frac{\partial E}{u_1} &= (z-r)h_1 \\
\frac{\partial E}{u_2} &= (z-r)h_2 \\
\frac{\partial E}{b_z} &= (z-r) \\
\frac{\partial E}{w_{11}} &= (z-r)u_1h_1(1-h_1)x_1 \\
\frac{\partial E}{w_{12}} &= (z-r)u_1h_1(1-h_1)x_2 \\
\frac{\partial E}{b_1} &= (z-r)u_1h_1(1-h_1) \\
\frac{\partial E}{w_{21}} &= (z-r)u_2h_2(1-h_2)x_1 \\
\frac{\partial E}{w_{22}} &= (z-r)u_2h_2(1-h_2)x_2 \\
\frac{\partial E}{b_2} &= (z-r)u_2h_2(1-h_2)
\end{aligned}$$

2.2. Compact Form.

$$\frac{\partial E}{\partial \mathbf{u}} = (z - r)\mathbf{h}^\top$$

$$\frac{\partial E}{\partial b_z} = (z - r)$$

$$\frac{\partial E}{\partial \mathbf{W}} = (z - r)((1 - \mathbf{h}) \odot \mathbf{h} \odot \mathbf{u}^\top) * \mathbf{x}^\top$$

$$\frac{\partial E}{\partial \mathbf{b}_h} = (z - r)((1 - \mathbf{h}) \odot \mathbf{h} \odot \mathbf{u}^\top)$$

3. EXPERIMENTS

The affect of hidden unit size is indicated in figure 1. More hidden units enables fast converges as using more parameters makes the problem more flexible. Combining less accurate planes gives results rapidly whereas it must be too much accurate with 2 hidden units that takes more time to be learned.

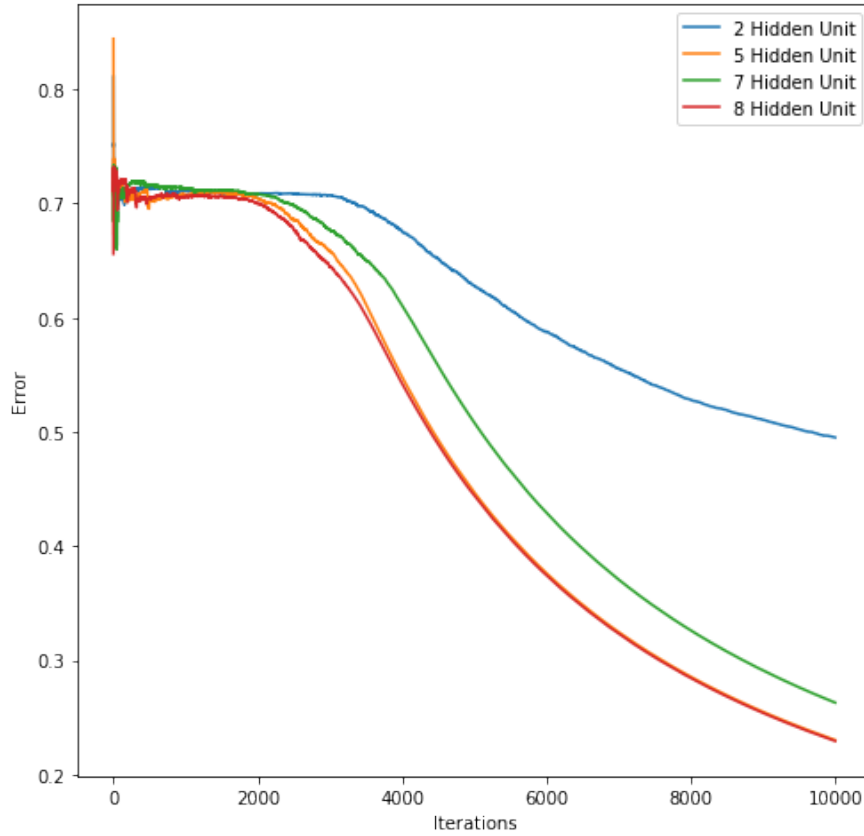
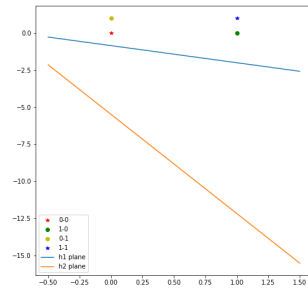
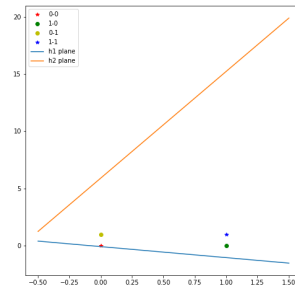


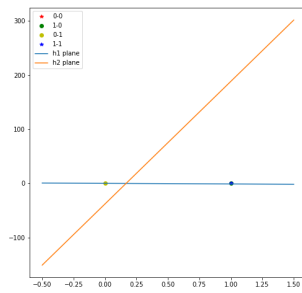
FIGURE 1. After 2000 iteration



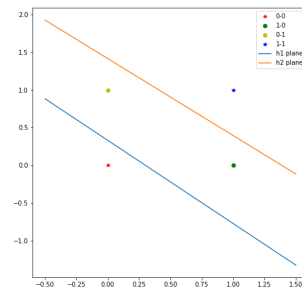
(A) After 1000 iteration



(B) After 2000 iteration

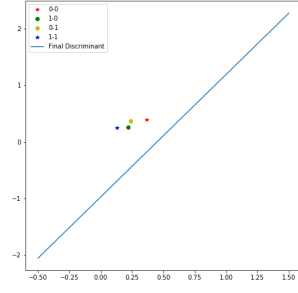


(C) After 3000 iteration

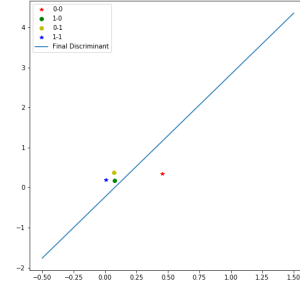


(D) After 4000 iteration

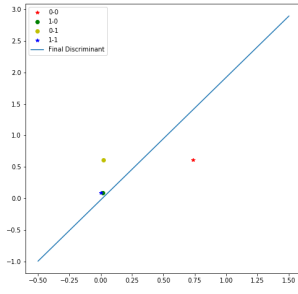
FIGURE 2. Evolution of Planes which generates hidden units



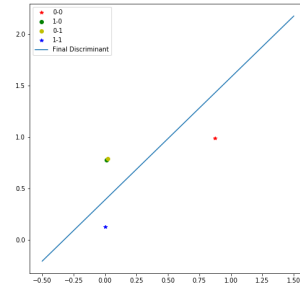
(A) After 1000 iteration



(B) After 2000 iteration



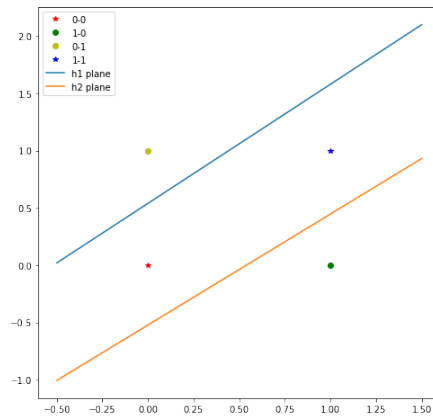
(C) After 3000 iteration



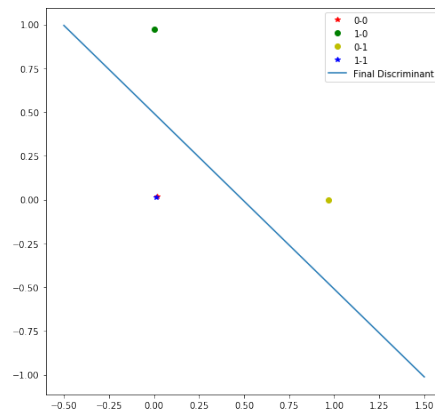
(D) After 4000 iteration

FIGURE 3. Evolution of Final Discriminant

4. ADDITIONAL MATERIAL



(A) Planes that generates hidden features



(B) Final Plane

```

: print('W_z: ',xor.W_z)
  print('b_z: ',xor.b_z)
  print('W_h1',xor.W_h[0,:])
  print('b_h1:',xor.b_h[0,0])
  print('W_h2',xor.W_h[1,:])
  print('b_h2:',xor.b_h[1,0])

W_z: [[ 14.80952344  14.76101927]]
b_z: [[-7.27497885]]
W_h1 [-7.65041714  7.36125404]
b_h1: -3.97000899784
W_h2 [ 7.64177225 -7.8921135 ]
b_h2: -4.12111727682

```

(C) Weights