1.

(a) NAND of two binary inputs

bias: 3

$X_1 \xrightarrow{-2} h \rightarrow \hat{y}$

$X_2 \xrightarrow{-2} h$
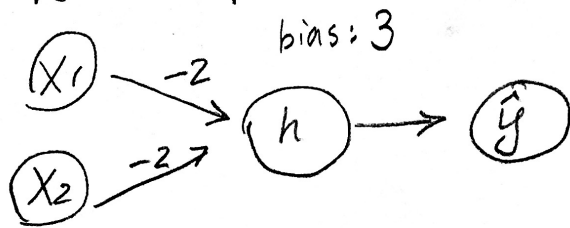
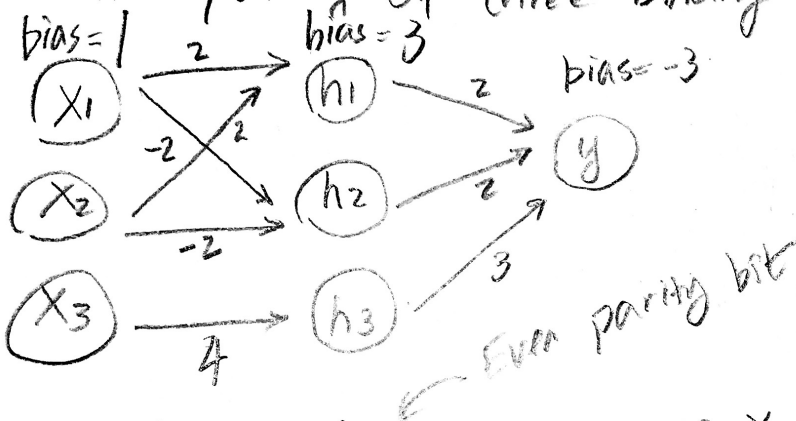| X_1	X_2	Output
0	0	1
0	1	1
1	0	1
1	1	0

$0 \cdot (-2) + 0 \cdot (-2) + 3 > 0 \Rightarrow 1$

$0 \cdot (-2) + 1 \cdot (-2) + 3 > 0 \Rightarrow 1$

$1 \cdot (-2) + 0 \cdot (-2) + 3 > 0 \Rightarrow 1$

$1 \cdot (-2) + 1 \cdot (-2) + 3 < 0 \Rightarrow 0$

(b) The parity of three binary inputs

bias=1, bias=3, bias=-3

$X_1, X_2, X_3 \rightarrow h_1, h_2, h_3 \rightarrow y$

weights: 2, -2, 2, -2, 4, 2, 2, 3

← even parity bit

X_1	X_2	X_3	Y
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
0	0	1	1
1	1	0	0
1	0	0	1
1	1	1	1

X_1 XOR X_2 XOR X_3

If X_3 is 0, it doesn't change the result of X_1 XOR X_2.

If X_3 is 1, it changes the result of X_1 XOR X_2

Extra Credit:

3. TANH results in better back propagation because TANH gives stronger gradients between value (-1, 1) while sigmoid gives a value between (0, 1).

Part II

1. Text Classification

table of performance statistics

```
+---------+------------------------+------------------------+
|Classifier|      Naives Bayes      |  SVM with Cosine Kernel |
+---------+------------------------+------------------------+
|         |    Train  |    Test    |    Train   |    Test    |
+---------+-----------+------------+-----------+-----------+
| Accuracy | 93.96 %  | 78.62 %    | 99.65 %   | 84.47 %    |
+---------+-----------+------------+-----------+-----------+
| Precision| 95.09 %  | 83.54 %    | 99.65 %   | 84.60 %    |
+---------+-----------+------------+-----------+-----------+
| Recall   | 92.64 %  | 76.85 %    | 99.64 %   | 83.75 %    |
+---------+-----------+------------+-----------+-----------+
|Train Time|      0.16 s           |        42.99 s          |
+---------+------------------------+------------------------+
```

SVM is better since it has better train and test accuracy even though it takes more time to train, and Naives Bays has the independence assumption which is not suitable for this text classification case.

2. Neural Network

Learning rate: 1.8

Regularization lambda: 0.0001

Best training accuracy: 96.24