Adaptive Computation & Machine Learning
Assignment #3
Due Date: 8:00PM, 22 March 2018

Question 1 K-Nearest Neighbour Classification {40 Marks}

Consider the following training and testing data sets where x_is are features and ys are class labels. For each record, features are extracted from images of genuine (-1) and forged (+1) banknote-like specimens. Soft copies of these data sets (`banknote_training_data.txt` and `banknote_testing_data.txt` respectively for training and testing data sets) are available under `Resources\data\assignment#3\` on Sakai.

Training Data Set

Record #	x_1	x_2	x_3	x_4	y
1	4.845100	8.111600	-2.951200	-1.472400	-1
2	1.311400	4.546200	2.293500	0.225410	-1
3	0.117390	6.276100	-1.549500	-2.474600	-1
4	0.774450	9.055200	-2.408900	-1.388400	-1
5	-0.166820	5.897400	0.498390	-0.700440	-1
6	2.046600	2.030000	2.176100	-0.083634	-1
7	4.002600	-3.594300	3.557300	0.268090	-1
8	3.137700	-4.109600	4.570100	0.989630	-1
9	-0.216610	8.032900	1.884800	-3.885300	-1
10	-0.398160	5.978100	1.391200	-1.162100	-1
11	0.740540	0.366250	2.199200	0.484030	-1
12	1.018200	9.109000	-0.620640	-1.712900	-1
13	2.506800	1.158800	3.924900	0.125850	-1
14	0.680870	2.325900	4.908500	0.549980	-1
15	4.926400	5.496000	-2.477400	-0.506480	-1
16	-1.966700	11.805200	-0.404720	-7.871900	-1
17	4.154200	7.275600	-2.476600	-1.209900	-1
18	-0.106480	-0.767710	7.757500	0.641790	-1
19	3.108800	3.112200	0.808570	0.433600	-1
20	3.899900	1.734000	1.601100	0.967650	-1
21	-1.670600	-2.090000	1.584000	0.711620	1
22	-1.693600	2.785200	-2.183500	-1.927600	1
23	-1.998300	-6.607200	4.825400	-0.419840	1
24	-4.002500	-13.497900	17.677200	-3.320200	1
25	-0.898090	-4.486200	2.200900	0.507310	1
26	-3.563700	-8.382700	12.393000	-1.282300	1
27	-1.663700	3.288100	-2.270100	-2.222400	1
28	0.266370	0.732520	-0.678910	0.035330	1
29	-3.571300	-12.492200	14.888100	-0.470270	1
30	-4.140900	3.461900	-0.478410	-3.887900	1
31	-2.482400	-7.304600	6.839000	-0.590530	1
32	-1.651400	-8.498500	9.112200	1.237900	1
33	-2.591200	-0.105540	1.279800	1.041400	1
34	-0.620430	0.558700	-0.385870	-0.664230	1
35	-0.873400	1.653300	-2.196400	-0.780610	1
36	-1.862900	-0.848410	2.537700	0.097399	1
37	-4.577000	3.451500	0.667190	-0.947420	1
38	-3.551000	1.895500	0.186500	-2.440900	1
39	-1.855400	-9.603500	7.776400	-0.977160	1
40	-2.799000	1.967900	-0.423570	-2.112500	1

Testing Data Set

x_1	x_2	x_3	x_4	y
1.781900	6.917600	-1.274400	-1.575900	-1
3.618100	-3.745400	2.827300	-0.712080	-1
-1.257600	1.589200	7.007800	0.424550	-1
4.743200	2.108600	0.136800	1.654300	-1
2.152600	-6.166500	8.083100	-0.343550	-1
3.543800	1.239500	1.997000	2.154700	-1
2.259600	-0.033118	4.735500	-0.277600	-1
4.289900	9.181400	-4.606700	-4.326300	-1
2.292800	9.038600	-3.241700	-1.299100	-1
4.336500	-3.584000	3.688400	0.749120	-1
2.392500	9.798000	-3.036100	-2.822400	-1
1.363800	-4.775900	8.418200	-1.883600	-1
3.971900	1.036700	0.759730	1.001300	-1
5.745600	10.180800	-4.785700	-4.336600	-1
0.324440	10.067000	-1.198200	-4.128400	-1
2.974200	8.960000	-2.902400	-1.037900	-1
-2.741900	11.403800	2.539400	-5.579300	-1
0.248350	7.643900	0.988500	-0.873710	-1
0.379800	0.709800	0.757200	-0.444400	-1
2.188100	2.735600	1.327800	-0.183200	-1
-3.580100	-12.930900	13.177900	-2.567700	1
-2.436500	3.602600	-1.416600	-2.894800	1
1.345100	0.235890	-1.878500	1.325800	1
-4.209100	4.728300	-0.491260	-5.215900	1
-0.556480	3.213600	-3.308500	-2.796500	1
-4.366700	6.069200	0.572080	-5.466800	1
-1.835600	-6.756200	5.058500	-0.550440	1
-2.620000	-6.855500	6.216900	-0.622850	1
0.539360	3.894400	-4.816600	-4.341800	1
-2.291800	-7.257000	7.959700	0.921100	1
1.227900	4.030900	-4.643500	-3.912500	1
-0.735100	1.736100	-1.493800	-1.158200	1
-2.314700	3.666800	-0.696900	-1.247400	1
-3.724400	1.903700	-0.035421	-2.509500	1
-2.790800	-5.713300	5.953000	0.459460	1
-2.909800	-10.071200	8.415600	-1.994800	1
-3.019300	1.777500	0.737450	-0.453460	1
-1.130600	1.845800	-1.357500	-1.380600	1
-4.553100	-12.585400	15.441700	-1.498300	1
-1.710100	-8.790300	7.973500	-0.454750	1

Table 1: Banknote Authentication Data Set

For the following questions on K-Nearest Neighbour (K-NN) classifier, complete the tables to show your calculations in terms of closest record number(s) from training data set, corresponding distance(s) and label(s). The final output of the K-NN will be entered in Prediction section. The records should be sorted in ascending order of their distances from the query data point. You will use Euclidean distance in search for nearest neighbours. Examples for the prediction of 1-NN and 3-NN classifications for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1.781900 \\ 6.917600 \\ -1.274400 \\ -1.575900 \end{bmatrix}$ are given below.

1-NN

Record #	Distance from the record	Label of the record
3	2.016298	-1
Prediction		-1

3-NN

Record #	Distance from the record	Label of the record
3	2.016298	-1
12	2.414879	-1
4	2.628029	-1
Prediction		-1

Q.1.1{5 Marks} What is the prediction of the 1-NN classifier for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3.618100 \\ -3.745400 \\ 2.827300 \\ -0.712080 \end{bmatrix}$?

Record #	Distance from the record	Label of the record
Prediction		

ANSWER:

Record #	Distance from the record	Label of the record
7	1.290079	-1
Prediction		-1

Q.1.2{5 Marks} What is the prediction of the 3-NN classifier for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2.436500 \\ 3.602600 \\ -1.416600 \\ -2.894800 \end{bmatrix}$?

Record #	Distance from the record	Label of the record
Prediction		

ANSWER:

Record #	Distance from the record	Label of the record
27	1.369932	1
22	1.656398	1
40	2.098035	1
Prediction		1

Q.1.3{5 Marks} What is the prediction of the 5-NN classifier for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -4.366700 \\ 6.069200 \\ 0.572080 \\ -5.466800 \end{bmatrix}$?

Record #	Distance from the record	Label of the record
Prediction		

ANSWER:

Record #	Distance from the record	Label of the record
30	3.231943	1
9	5.030282	-1
37	5.227851	1
38	5.233534	1
40	5.614356	1
Prediction		1

Q.1.4{25 Marks} Using the above training and testing data sets given in Table 1, calculate test error rates for 3-NN and 5-NN classifiers to compete the following table where test error rate for 1-NN classifier is provided for you.

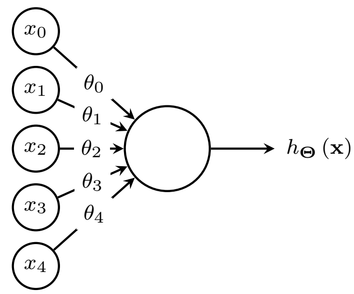
1-NN Test Error Rate	3-NN Test Error Rate	5-NN Test Error Rate
0.025		

ANSWER:

1-NN Test Error Rate	3-NN Test Error Rate	5-NN Test Error Rate
0.025	0.050	0.075

Question 2 Logistic Regression {40 Marks}

Consider the following logistic regression model where $x_0 = 1.0$ and x_1, x_2, x_3 and x_4 are input features as defined in Table 1.



Let $\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^5$ be the input feature vector, $\mathbf{\Theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \in \mathbb{R}^5$ be the parameter vector, and

$$h_{\mathbf{\Theta}}(\mathbf{x}) = 2\frac{1}{1+e^{-\mathbf{\Theta}^T\mathbf{x}}} - 1 = 2\frac{1}{1+e^{-(\theta_0 x_0+\theta_1 x_1+\theta_2 x_2+\theta_3 x_3+\theta_4 x_4)}} - 1$$

be the output of the logistic regression model.

Q.2.1{5 Marks} The range of $g(z) = \frac{1}{1+e^{-z}}$ is $[0, 1]$. What is the range of $h_{\mathbf{\Theta}}(\mathbf{x})$?

ANSWER:

The range of $h_{\mathbf{\Theta}}(\mathbf{x})$ is $2[0, 1] - 1 = [0, 2] - 1 = [-1, 1]$.

Q.2.2{20 Marks} For a given training set $\left\{\left(\mathbf{x}^{(n)}, y^{(n)}\right); n = 1, \ldots, N\right\}$,$\mathbf{x}^{(n)} = \begin{bmatrix} x_0^{(n)} \\ x_1^{(n)} \\ x_2^{(n)} \\ x_3^{(n)} \\ x_4^{(n)} \end{bmatrix} \in \mathbb{R}^5$, $y^{(n)} \in \{-1, 1\}$, and cost function $J(\mathbf{\Theta}) = \frac{1}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)^2$, derive closed form parameter update rules for θ_k using batch gradient descent algorithm, i.e.,

$$\theta_k \leftarrow \theta_k - \alpha\frac{\partial J(\mathbf{\Theta})}{\partial \theta_k}, \forall k = \{0, \ldots, 4\}.$$

ANSWER:

$$\begin{aligned}
\frac{\partial J(\mathbf{\Theta})}{\partial \theta_k} &= \frac{\partial}{\partial \theta_k}\frac{1}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)^2 \\
&= \frac{2}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)\frac{\partial}{\partial \theta_k}h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) \\
&= \frac{2}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)\frac{\left(1+h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)\left(1-h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)}{2}\mathbf{x}_k^{(n)} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)\left(1 - h_{\mathbf{\Theta}}^2\left(\mathbf{x}^{(n)}\right)\right)\mathbf{x}_k^{(n)}
\end{aligned}$$

Thus,

$$\theta_k^{(t)} = \theta_k^{(t-1)} - \alpha\frac{1}{N}\sum_{n=1}^{N}\left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)\left(1 - h_{\mathbf{\Theta}}^2\left(\mathbf{x}^{(n)}\right)\right)\mathbf{x}_k^{(n)}.$$

Q.2.3{15 Marks} Suppose that $\mathbf{\Theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \begin{bmatrix} 1.47791190 \\ -1.55084741 \\ -0.89930302 \\ -0.91395829 \\ -0.14429827 \end{bmatrix}$ is obtained after running batch gradient descent learning algorithm. The output of logistic regression model for an input $\mathbf{x}^{(n)}$ is mapped to a binary output (1 or -1) according to the following equation

$$\begin{array}{rl} 1 & \text{for} \quad h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) > 0; \\ -1 & \text{for} \quad h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) \leq 0. \end{array}$$

Given the above information and the testing data in Table 1, complete the following table

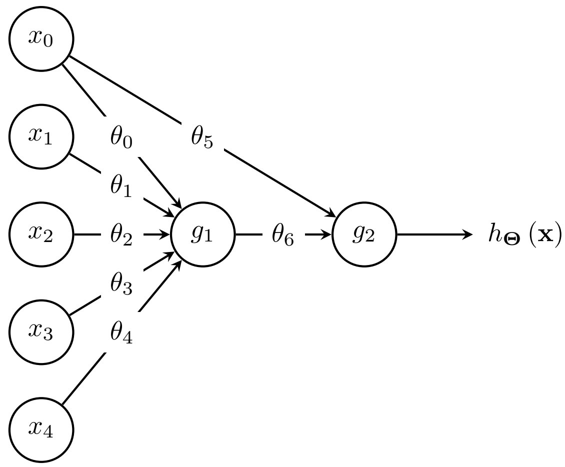
Testing record #	Classifier Output $h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)$	Final Output
5	-0.949479	-1
10		
15		
20		
25		
30		
35		
40		

ANSWER:

Testing record #	Classifier Output $h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)$	Final Output
5	-0.949479	-1
10	-0.991887	-1
15	-0.996641	-1
20	-0.992353	-1
25	0.893525	1
30	0.968980	1
35	0.991332	1
40	0.983889	1

Question 3 Artificial Neural Networks {20 Marks}

Consider the following 3-layer ANN model where $\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^5$ is input vector, θ_js are network parameters, g_ks are neuron activation functions, and $h_{\mathbf{\Theta}}(\mathbf{x})$ is the network output (or hypothesis).



Let $g_1(z) = g_2(z) = \frac{1}{1+e^{-z}}$ for the input $z \in \mathbb{R}$.
Q.3.1{10 Marks} What is the output $h_{\mathbf{\Theta}}(\mathbf{x})$ of the ANN in terms of x_is, θ_js and g_ks?

ANSWER:

$$h_{\mathbf{\Theta}}(\mathbf{x}) = g_2\left(\theta_5 x_0 + \theta_6 g_1\left(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4\right)\right)$$

Q.3.2{10 Marks} The parameters of $h_{\mathbf{\Theta}}(\mathbf{x})$ are learned using a training set $\left\{\left(\mathbf{x}^{(n)}, y^{(n)}\right)\right\}_{n=1}^{N}$ of size N, where each training sample $\left(\mathbf{x}^{(n)}, y^{(n)}\right)$ is a tuple of a feature vector $\mathbf{x}^{(n)}$ with a label $y^{(n)}$ of positive class ($y^{(n)} = 1$) or negative class ($y^{(n)} = 0$). The parameters are iteratively learned by minimizing a cost function ($J(\mathbf{\Theta})$) between the real labels $y^{(n)}$s and the predicted labels $h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)$s resulted from the hypothesis evaluation, i.e.,

$$J(\mathbf{\Theta}) = \frac{1}{2N} \sum_{n=1}^{N} \left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)^2.$$

As a minimizer of the cost function, gradient descent algorithm with a learning rate $\alpha \in (0, 1]$ can be employed to obtain iterative minimizers according to

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial J(\mathbf{\Theta})}{\partial \theta_j}.$$

Derive closed form equations of iterative minimizers for all network parameters θ_js.

ANSWER:
Let $s_1 = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$, $u_1 = g_1(s_1)$, $s_2 = \theta_5 x_0 + \theta_6 u_1$, $h_{\mathbf{\Theta}}(\mathbf{x}) = u_2 = g_2(s_2)$

$$\begin{aligned}
\frac{\partial J(\mathbf{\Theta})}{\partial \theta_6} &= \frac{1}{N} \sum_{n=1}^{N} \left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right) \frac{\partial h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_6} \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right) u_1\left(\mathbf{x}^{(n)}\right) h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) \left(1 - h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J(\mathbf{\Theta})}{\partial \theta_5} &= \frac{1}{N} \sum_{n=1}^{N} \left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right) \frac{\partial h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_5} \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right) x_0 h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right) \left(1 - h_{\mathbf{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\Theta})}{\partial \theta_4} &= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)\frac{\partial h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_4} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)x_4 u_1\left(\mathbf{x}^{(n)}\right)\left(1-u_1\left(\mathbf{x}^{(n)}\right)\right)\theta_6 h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\left(1-h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\Theta})}{\partial \theta_3} &= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)\frac{\partial h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_3} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)x_3 u_1\left(\mathbf{x}^{(n)}\right)\left(1-u_1\left(\mathbf{x}^{(n)}\right)\right)\theta_6 h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\left(1-h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\Theta})}{\partial \theta_2} &= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)\frac{\partial h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_2} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)x_2 u_1\left(\mathbf{x}^{(n)}\right)\left(1-u_1\left(\mathbf{x}^{(n)}\right)\right)\theta_6 h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\left(1-h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\Theta})}{\partial \theta_1} &= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)\frac{\partial h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_1} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)x_1 u_1\left(\mathbf{x}^{(n)}\right)\left(1-u_1\left(\mathbf{x}^{(n)}\right)\right)\theta_6 h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\left(1-h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\Theta})}{\partial \theta_0} &= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)\frac{\partial h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)}{\partial \theta_0} \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)-y^{(n)}\right)x_0 u_1\left(\mathbf{x}^{(n)}\right)\left(1-u_1\left(\mathbf{x}^{(n)}\right)\right)\theta_6 h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\left(1-h_{\boldsymbol{\Theta}}\left(\mathbf{x}^{(n)}\right)\right)
\end{aligned}
$$