Problem Set 1
CSCI 5992 - Neural Networks and Deep Learning

Rajeev R Menon - 110581437

November 3, 2022

Problem 1

Artificial intelligence is a programmable system that can think, act and learn like a regular human brain. It is a machine that can simulate human intelligence. It is used for automating and solving various problems which otherwise require human intelligence. This is achieved using various algorithms. These machines take input in various format, which is then put through the algorithm to produce the desired output.

Machine learning is a subset of artificial intelligence that deals with the algorithms that are required to analyze and learn from data. This knowledge is then used to make predictions to solve various real-world problems. There are various categories of machine learning algorithms like supervised learning, unsupervised learning, etc.

Deep learning is a subset of machine learning which uses algorithms that are inspired by the functioning of a human brain. Just like a human brain, deep learning uses multiple layers of artificial neurons to process data and make decisions. This enables it to process large amounts of data similar to an actual human brain. Unlike other machine learning algorithms, deep learning doesn't require any specification of existing knowledge to make decisions, but rather it learns from experience, much similar to an actual human.

Problem 2

Part a

A training dataset is a set of labeled data used to train the model. The model is trained multiple times over the training data to learn more about the data and adjust the weights and biases. The learning obtained from the testing dataset is used later to make predictions on data the model hasn't seen before.

After the model has been trained using the training set, it is run through a testing dataset to predict the output. This is a simulation of the real-world scenario when the model is deployed for use in production. The performance of the model on the testing dataset can be used to evaluate and further optimize the model as required.

Part b

There shouldn't be any overlap of data between training and test datasets. After multiple epochs of learning over the training dataset, the model becomes good at making predictions from the training dataset. Any overlap of data between training and test datasets would result in a high performance of the model over the testing dataset as well. This is not a good indicator of real-world performance when the model is fed data that it hasn't seen before.

Problem 3

Part a

Training data:

Sample	X1	X2	X3	Y
1	1	0	1	1
2	1	1	0	1
3	1	0	0	-1

Learning rate:

$$\eta = 0.1$$

Initial weights

w_0	w_1	w_2	w_3
0	0	0	0

$$x_0 = 1$$

Prediction:

$$\sum_{j=0}^{m} x_j w_j = w^T x$$

$$\phi(w^T x) = \begin{cases} 1 & w^T x \geq 0 \\ -1 & otherwise \end{cases}$$

Change in weights:

$$\Delta w_j = \eta(target^{(i)} - output^{(i)})x_j^{(i)}$$

Epoch 1:

Predicted value for sample 1

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * 0 + 1 * 0 + 0 * 0 + 1 * 0 \\ &= 0 \\ \phi(w^T x) &= 1 \end{aligned}$$

Change in weights:

$$\begin{aligned} \Delta w_0 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_1 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_2 &= 0 + 0.1 * (1 - 1) * 0 = 0 \\ \Delta w_3 &= 0 + 0.1 * (1 - 1) * 1 = 0 \end{aligned}$$

Updated model:

Sample	X1	X2	X3	X4
1	1	0	1	1
2	1	1	0	1

Predicted
1
?

w0	w1	w2	w3
0	0	0	0
0	0	0	0

Predicted value for sample 2:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * 0 + 1 * 0 + 1 * 0 + 0 * 0 \\ &= 0 \\ \phi(w^T x) &= 1 \end{aligned}$$

Change in weights:

$$\begin{aligned} \Delta w_0 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_1 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_2 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_3 &= 0 + 0.1 * (1 - 1) * 0 = 0 \end{aligned}$$

Updated Model:

Sample	X1	X2	X3	X4
1	1	0	1	1
2	1	1	0	1
3	1	0	0	-1

Predicted
1
1
?

w0	w1	w2	w3
0	0	0	0
0	0	0	0
0	0	0	0

Predicted value for sample 3

$$\begin{aligned}
w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\
&= 1 * 0 + 1 * 0 + 0 * 0 + 0 * 0 \\
&= 0 \\
\phi(w^T x) &= 1
\end{aligned}$$

Updated weights:

$$\begin{aligned}
\Delta w_0 &= 0 + 0.1 * (-1 - 1) * 1 = -0.2 \\
\Delta w_1 &= 0 + 0.1 * (-1 - 1) * 1 = -0.2 \\
\Delta w_2 &= 0 + 0.1 * (-1 - 1) * 0 = 0 \\
\Delta w_3 &= 0 + 0.1 * (-1 - 1) * 0 = 0
\end{aligned}$$

Updated weights after Epoch 1

w0	w1	w2	w3
0	0	0	0
0	0	0	0
0	0	0	0
-0.2	-0.2	0	0

(1)

Epoch 2:

Predicted value for sample 1

$$\begin{aligned}
w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\
&= 1 * -0.2 + 1 * -0.2 + 0 * 0 + 1 * 0 \\
&= -0.4 \\
\phi(w^T x) &= -1
\end{aligned}$$

Change in weight:

$$\begin{aligned}
\Delta w_0 &= -0.2 + 0.1 * (1 + 1) * 1 = 0 \\
\Delta w_1 &= -0.2 + 0.1 * (1 + 1) * 1 = 0 \\
\Delta w_2 &= 0 + 0.1 * (1 + 1) * 0 = 0 \\
\Delta w_3 &= 0 + 0.1 * (1 + 1) * 1 = 0.2
\end{aligned}$$

Updated model:

Sample	X1	X2	X3	X4
1	1	0	1	1
2	1	1	0	1

Predicted
-1
?

w0	w1	w2	w3
-0.2	-0.2	0	0
0	0	0	0.2

Predicted value for sample 2:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * 0 + 1 * 0 + 1 * 0 + 0 * 0.2 \\ &= 0 \\ \phi(w^T x) &= 1 \end{aligned}$$

Change in weights:

$$\begin{aligned} \Delta w_0 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_1 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_2 &= 0 + 0.1 * (1 - 1) * 1 = 0 \\ \Delta w_3 &= 0.2 + 0.1 * (1 - 1) * 0 = 0.2 \end{aligned}$$

Updated model

Sample	X1	X2	X3	X4
1	1	0	1	1
2	1	1	0	1
3	1	0	0	-1

Predicted
-1
1
?

w0	w1	w2	w3
-0.2	-0.2	0	0
0	0	0	0.2
0	0	0	0.2

Predicted value for sample 3:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * 0 + 1 * 0 + 0 * 0 + 0 * 0.2 \\ &= 0 \\ \phi(w^T x) &= 1 \end{aligned}$$

Change in weights:

$$\begin{aligned} \Delta w_0 &= 0 + 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_1 &= 0 + 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_2 &= 0 + 0.1 * (-1 - 1) * 0 = 0 \\ \Delta w_3 &= 0.2 + 0.1 * (-1 - 1) * 0 = 0.2 \end{aligned}$$

Final weights after 2 Epochs

Epoch	w0	w1	w2	w3
0	0	0	0	0
1	-0.2	-0.2	0	0
2	-0.2	-0.2	0	0.2

Part b

Predicted value for sample 1:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * -0.2 + 1 * -0.2 + 1 * 0 + 0 * 0.2 \\ &= -0.4 \\ \phi(w^T x) &= -1 \end{aligned}$$

Predicted value for sample 2:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * -0.2 + 1 * -0.2 + 0 * 0 + 1 * 0.2 \\ &= -0.2 \\ \phi(w^T x) &= -1 \end{aligned}$$

Predicted value for sample 3:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * -0.2 + 1 * -0.2 + 1 * 0 + 1 * 0.2 \\ &= -0.2 \\ \phi(w^T x) &= -1 \end{aligned}$$

Predicted value for sample 4:

$$\begin{aligned} w^T x &= x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \\ &= 1 * -0.2 + 0 * -0.2 + 0 * 0 + 0 * 0.2 \\ &= -0.2 \\ \phi(w^T x) &= -1 \end{aligned}$$

Part c

Confusion matrix:

		Predicted	
		Y=1	Y=-1
Actual	Y=1	0	2
	Y=-1	0	2

Part d

$$
\begin{aligned}
Accuracy &= \frac{TP + TN}{Total} \\
&= \frac{2}{4} \\
&= 0.5 \\
Precision &= \frac{TP}{TP + FP} \\
&= \frac{0}{0} \\
&= undefined \\
Recall &= \frac{TP}{TP + FN} \\
&= \frac{0}{2} \\
&= 0
\end{aligned}
$$