CSCI 5922 Fall 2022–HW1

Abulitibu Tuguluke

September 9, 2022

1 Artificial Intelligence versus Machine Learning versus Deep Learning

I think before we define what Artificial Intelligence is, we need to understand Intelligence first. It is an ability to acquire knowledge, or on a higher note, the ability to enjoy acquire and apply skills, or solve problems. By adding an adjective 'Artificial', we can say it is a machine or system's ability to acquire knowledge. Better yet, a computer enjoys playing chess without us telling it to.
If we followed the 3 nested diagram: (A.I.(ML(DL))), Machine learning seems to be the washed-down version of A.I. that is constrained by data, it is still about a machine's ability to learn/acquire their knowledge, yet with data extraction.
My first impression of Deep Learning was it is just Numerical linear Algebra(HPC) on 'steroid', by converting text/images into vector/matrix/tensors, and then feed them into GPUs, drink some coffee while waiting for them to be converted back to new text/images at the end. Later I realize it is more than that: it is still part of ML, a system that acquires knowledge through pattern extraction from data, but the processing has to be through representation learning(automatic feature extraction), with all kinds of neural network architecture. My guess the word deep came from multiple layers of network instead of a single.
To quote Judea Pearl's Ladder of Causation: Seeing $\rightarrow$ Doing $\rightarrow$ Imagining. [1]. I believe the difference between these three disciplines is A.I. is the ultimate ladder: be able to imagine and understand the counterfactual (things that are not in the raw data). Maching learning is the middle ladder: be able to do and intervene by leveraging raw data. Finally, Deep learning is the first ladder: observe and associate, through particular architectures and learning features. Albeit the differences, they are still the same ladder.

2 Supervised Learning Generalization

(a)

Imagine a scenario where you train all the raw data set in your model, to the point that there is no 'error' (the model getting all the label right in the training set), yet once we put it into production and predicting the 'real world' data set, and your product manager come back and accuse your model is getting the result wrong. There could be many reasons for that, the data itself is wrong, or unbalanced, but most likely due to over-fitting, the model did not 'learn' but 'memorize' the pattern from raw data, like a bad student. Hence, we need to introduce the train/test split, to always 'surprise' the student (your model) that there is always a space to learn, not just memorize.

(b)

The only case that I can think of for the overlapping is 'continuous' training the models, e,g,: a resume scanning service, where we continuously inject new resume into the training set after we got our performance result for the test set, by including some of it.
Otherwise, I would say no. Overlapping the training and test set may give you a boost in your ego for model performance, yet in real production, it may not be as good you think it is. So NO.

[1] The Book of Why: The New Science of Cause and Effect

3 Artificial Neurons

(a) Training

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

Epoch 1

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

Predicted
$\Phi(w^T x) = 1$ since $(\mathbf{w}^T \mathbf{x}) = 0$
$\Phi(w^T x) = 1$ since $(\mathbf{w}^T \mathbf{x}) = 0$
$\Phi(w^T x) = 1$ since $(\mathbf{w}^T \mathbf{x}) = 0$

w_0	w_1	w_2	w_3	Comment for weight update
0	0	0	0	No updates since prediction matches ground truth
0	0	0	0	No updates since prediction matches ground trut
0	0	0	0	Now we need update since prediction not matching ground truth:: 1
-0.2	-0.2	0	0	End of Epoch 1

$$\begin{cases} \Delta w_0 = 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_1 = 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_2 = 0.1 * (-1 - 1) * 10 = 0 \\ \Delta w_3 = 0.1 * (-1 - 1) * 0 = 0 \end{cases} \tag{1}$$

Epoch 2

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

Predicted
$\Phi(w^T x) = -1$ since $(\mathbf{w}^T \mathbf{x}) = 1 * -0.2 = -0.2$
$\Phi(w^T x) = 1$ since $(\mathbf{w}^T \mathbf{x}) = 1 * 0.2 = 0.2$
$\Phi(w^T x) = 1$ since $(\mathbf{w}^T \mathbf{x}) = 1 * 0.2 = 0.2$

w_0	w_1	w_2	w_3	Comment for weight update
-0.2	-0.2	0	0	Now we need update since prediction not matching ground truth: 2
0.2	0.2	0	0.2	No updates since prediction matches ground truth
0.2	0.2	0	0.2	Now we need update since prediction not matching ground truth: 3
-0.2	-0.2	0	0	End of Epoch 2

$$\begin{cases} \Delta w_0 = 0.1 * (1 - -1) * 1 = 0.2 \\ \Delta w_1 = 0.1 * (1 - -1) * 1 = 0.2 \\ \Delta w_2 = 0.1 * (1 - -1) * 0 = 0 \\ \Delta w_3 = 0.1 * (1 - -1) * 1 = 0.2 \end{cases} \tag{2}$$

$$\begin{cases} \Delta w_0 = 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_1 = 0.1 * (-1 - 1) * 1 = -0.2 \\ \Delta w_2 = 0.1 * (-1 - 1) * 0 = 0 \\ \Delta w_3 = 0.1 * (-1 - 1) * 0 = 0 \end{cases} \tag{3}$$

(b) Test

X_1	X_2	X_3	Y	Prediction
1	1	0	-1	-1 since $(\mathbf{w}^T \mathbf{x}) = 1 * -0.2 + 1 * -0,2 = -0.4$
1	0	1	-1	-1 since $(\mathbf{w}^T \mathbf{x}) = 1 * -0.2- = -0.2$
1	1	1	1	-1 since $(\mathbf{w}^T \mathbf{x}) = 1 * -0.2 + 1 * -0,2 = -0.4$
0	0	0	1	1 since $(\mathbf{w}^T \mathbf{x}) = 0$

Table 1: Cofusion matrix

		Actual	
		1	-1
Predicted	1	1	0
	-1	1	2

(c)

(d)

$$\begin{cases} \text{Precision} = \dfrac{TP}{TP+FP} = \dfrac{1}{1+0} = 1 \\ \\ \text{Recall} = \dfrac{TP}{TP+FN} = \dfrac{1}{1+1} = 50\% \\ \\ \text{Accuracy} = \dfrac{TP+TN}{Total} = \dfrac{1+2}{1+2+1} = 75\% \end{cases}$$