
Assignment: Chapter 6: A Closer Look at Machines That Learn

1. TRUE/FALSE - The learning-from-data approach of deep neural networks has generally proved to be more successful than the “good old-fashioned AI” strategy, in which human programmers construct explicit rules for intelligent behavior. However, contrary to what some media have reported, the learning process of ConvNets is not very humanlike.

1A. True

2. Why does your professor like the previous question?

2A. ConvNets are not human like learning and the “good old-fashion AI” strategy is more interesting as a machine can give reasoning behind it’s decision whereas ConvNets cannot.

3. TRUE/FALSE - As we’ve seen, the most successful ConvNets learn via a supervised-learning procedure: they gradually change their weights as they process the examples in the training set again and again, over many epochs (that is, many passes through the training set), learning to classify each input as one of a fixed set of possible output categories.

3A. True

4. List some significant differences between the way that humans learn about objects and the way that ConvNets learn about objects.

4A. Moreover, children don’t learn passively: they ask questions, they demand information on the things they are curious about, they infer abstractions of and connections between concepts, and above all, they explore the world.

5. Why is it inaccurate to say that today’s successful ConvNets “learn on their own?”

5A. In order for a ConvNet to learn to perform a task, a huge amount of human effort is required to collect, curate, and label the data, as well as to design the many aspects of the ConvNet’s architecture.

6. In answer to the rhetorical question “Where does all of the data come from to fuel big data applications?,” MM answers “You - and probably everyone you know.” Please elaborate on the answer.

6A. Modern computer-vision applications are possible only because of the billions of images that internet users have uploaded and (sometimes) tagged with text identifying what is in the image.

7. How do car companies acquire the big data (labelled images of pedestrians, cyclists and other obstacles) needed to train robocars?

7A. Self-driving car companies collect these training examples from countless hours of video taken by cameras mounted on actual cars driving in traffic on highways and city streets.

8. What is the “long tail” phenomenon, and how does it relate to machines that learn (ConvNets)?

8A. This issue is compounded by the so-called long tail problem: the vast range of possible unexpected situations an AI system could be faced with. But once self-driving cars are widespread, while each individual unlikely situation is by definition, very unlikely, there are so many possible scenarios in the world of driving and so many cars that some self-driving car somewhere is likely to encounter one of them at some point.

9. TRUE/FALSE - A commonly proposed solution to the long tail problem in AI systems is to complement supervised learning with unsupervised learning.

9A. True

10. What is “unsupervised learning?”

10A. The term unsupervised learning refers to a broad group of methods for learning categories or actions without labeled data.

11. What colorful remark did Yann LeCun make about unsupervised learning?

11A. “unsupervised learning is the dark matter of AI”

12. TRUE/FALSE - For general AI, almost all learning will have to be unsupervised, but no one has yet come up with the kinds of algorithms needed to perform successful unsupervised learning.

12A. True

13. TRUE/FALSE - Humans have a fundamental competence lacking in current AI systems: common sense. We have vast background knowledge of the world, both its physical and social aspects. We have a good sense of how objects - both animate and living - are likely to behave, and we use this knowledge extensively in making decisions about how to act in any given situation.

13A. True

14. TRUE/FALSE - Many people believe that until AI systems have common sense as

humans do, we won't be able to trust them to be fully autonomous in complex real-world situations.

14A. True

15. TRUE/FALSE - Superficial changes to images, such as slightly blurring or speckling an image, changing some colors, or rotating objects in the scene, can cause ConvNets to make significant errors even when these perturbations don't affect humans' recognition of objects. This unexpected fragility of ConvNets – even those that have been said to “surpass humans at object recognition” – indicates that they are overfitting to their training data and learning something different from what we are trying to teach them, a phenomenon that results in various manifestations of unreliability.

15A. True

16. The unreliability of ConvNets can result in embarrassing – and potentially damaging – errors. Select a particularly embarrassing/damaging example of unreliability in ConvNets, and describe it in just a sentence or two.

16A. Google suffered a public relations nightmare in 2015 after it rolled out an automated photo-tagging feature (using a ConvNet) in its Photos app. In addition to correctly tagging images with generic descriptions such as “Airplanes,” “Cars,” and “Graduation,” the neural network tagged a selfie featuring two African Americans as “Gorillas,”

17. At the end of the section on biased AI, MM observes that the problem of bias in applications of AI has been getting a lot of attention recently, with many articles, workshops, and even academic research institutes devoted to this topic. What questions does she raise in conjunction with this observation? What do you think are the appropriate answers to these questions?

17A. Should the data sets being used to train AI accurately mirror our own biased society – as they often do now – or should they be tinkered with specifically to achieve social reform aims? And who should be allowed to specify the aims or do the tinkering?

17A pt 2. If we are to create a true “AI” should we chain the way the AI should think and use it as means to further human ambitions when such ambitions change from person to person. What are we creating AI for? When this question is answered those other questions can be answered.

18. TRUE/FALSE - You can often trust that people know what they are doing if they can explain to you how they arrived at an answer or a decision. However, “showing their work” is something that deep neural networks – the bedrock of AI systems – cannot easily do.

18A. True

19. TRUE/FALSE - Recall that a convolutional neural network decides what object is contained in an input image by performing a sequence of mathematical operations (convolutions) propagated through many layers. For a reasonably sized network, these can amount to billions of arithmetic operations. While it would be easy to program the computer to print out a list of all the additions and multiplications performed by a network for a given input, such a list would give us humans zero insight into how the network arrived at its answer. A list of a billion operations is not an explanation that a human can understand.

19A. True

20. What, according to MIT's Technology Review is the dark secret at the heart of AI?

20A. The fear is that if we don't understand how AI systems work, we can't really trust them or predict the circumstances under which they will make errors.

21. What does the phrase "theory of mind" refer to, and how is it related to our interactions with AI systems such as deep networks?

21A. A model of the other person's knowledge and goals in particular situations. None of us have a similar "theory of mind" for AI systems such as deep networks, which makes it harder to trust them.

22. One of the hottest new areas of AI is variously called "explainable AI," "transparent AI," or "interpretable machine learning." To what do these terms refer?

22A. These terms refer to research on getting AI systems – particularly deep networks – to explain their decisions in a way that humans can understand.

23. The field of "adversarial learning" has emerged in response to the fact that AI systems can readily be fooled in dramatic fashion, like mixing up a guy in glasses with Milla Jovovich, or misclassifying a stop sign for a speed-limit sign. Briefly describe the field of adversarial learning.

23A. Developing strategies that defend against potential (human) adversaries who could attack machine-learning systems.

24. Jeff Clune, an AI researcher at the University of Wyoming, made a very provocative analogy when he noted that there is "a lot of interest in whether Deep Learning is 'real intelligence' or a 'Clever Hans.'" Explain the essential question that underlies this analogy, being sure to incorporate a few words on the actual Clever Hans.

24A. Clever Hans became an international celebrity, a careful investigation eventually revealed that the horse did not actually understand the question or mathematical concepts put to him, but was tapping in response to subtle, unconscious cues given by

the questioner. Does deep learning exhibit “true understanding.” Or is it instead a computation Clever Hans responding to superficial cues in the data?