Chapter 6: Solution Document

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.

True

2. Why does your professor like the previous question?

Professor Graci likes the previous question because it suggests that symbolic AI is true artificial intelligence.

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.

True

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

ConvNet trains with data training sets and adjusting the weights. Humans are active learners who ask questions and explore the answers. Humans also do not need huge training sets in order to categorize objects.

5. Why is it inaccurate to say that today's successful ConvNets "learn on their own?"

Humans are the ones tuning the hyperparameters, also humans are the ones who create the training sets using human made data

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.

MM means that the data that fuels these applications come from interacting with the software and applications that collect the data. One example would be forms of social media whenever a photo is posted with you in it.

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

The companies collect training examples from videos taken from cameras mounted on cars that are driving in traffic.

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

The "long tail" phenomenon accounts for the reason why it is so difficult to train AI systems for self-driving cars. Essentially, there are so many possibilities for what can happen on the road – some common, many uncommon.

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

True

10. What is "unsupervised learning?"

Unsupervised learning is a group of methods for machine learning of categories and actions using labeled data.

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

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.

True

13. TRUE/FALSE - Humans have a fundamental competence lacking in current AI sysgtems: 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.

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.

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.

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 de- scribe it in just a sentence or two.

A image tagging system used by Google label a photo of 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?

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.

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.

True

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

Even the humans who train deep networks generally cannot look under the hood and provide explanation for the decision their networks make.

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?

The theory of mind is a model of the other person's knowledge and goal in particular situations.

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?

They refer to the field of AI that is trying to create deep learning systems that can explain its behavior.

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.

Adversarial learning is studying weaknesses in machines and providing solutions in order to take protective measures against malicious attacks on 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.