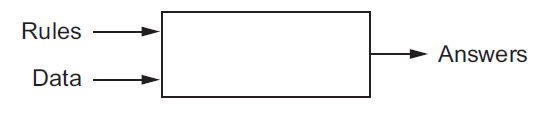
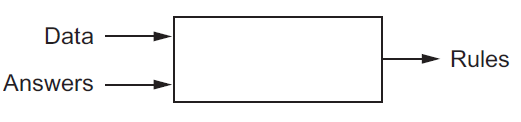
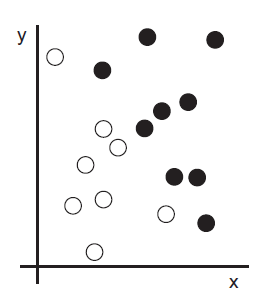
**Chapter 1 Quiz**

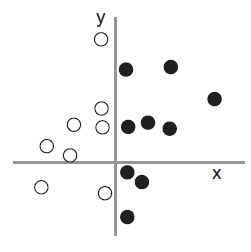
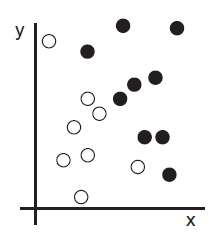
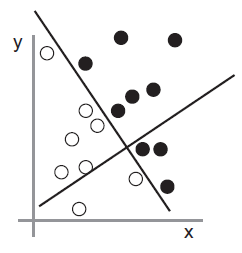
*What is deep learning?*

1. Place ‘*classical programming*’ and ‘*machine learning*’ in the correct empty boxes?
   1. 
   2. 
2. Is a machine-learning system explicitly programmed? (True or False)
3. Who is considered the pioneer of AI and what year was his landmark paper “Computing Machinery and Intelligence” published?
   1. Alan Greenspan 1955
   2. Bill Gates 1985
   3. Alan Turing 1950
   4. Alan Turing 1980
4. When is machine learning more efficient than Bayesian analysis?
   1. Large and complex datasets
   2. Classical statistical analysis
   3. Understanding linear algebra
   4. Binomial inefficiencies
5. Put the following deep learning steps in order.
   1. Examples of the expected output
   2. A way to measure whether the algorithm is doing a good job
   3. Input data points

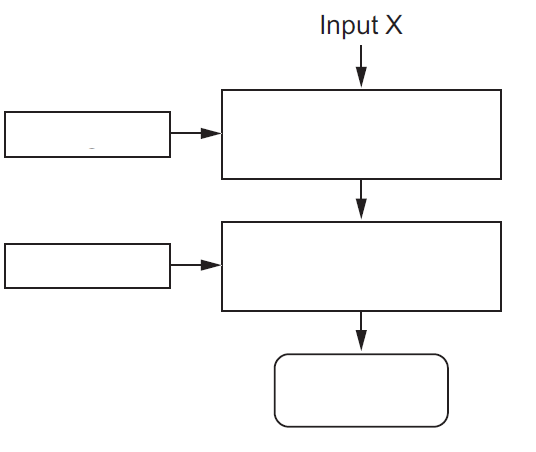
\_\_\_, \_\_\_, \_\_\_

1. Please select the best way to represent the sample data



* 1. 
  2. 
  3. 
  4. None

1. What is “*learning*” in machine learning?
   1. A process to create methods of solutions
   2. An automatic search process for better representations
   3. To create new data points
   4. A search to find outliers in data sets
2. What is a ‘hypothesis space’?
   1. A predefined set of hypotheses
   2. An area on a plane
   3. Space to write down hypotheses
   4. Hypotheses on a plane
3. What are neural networks?
   1. Networks on an electrical grid
   2. Networks that connecting a robotic limb
   3. Layered representations in deep learning
   4. Connected dots on a neural axis
4. What is the difference between shallow learning and deep learning?
5. How is deep learning like a water filter?
6. Fill in the empty boxes with the following words.

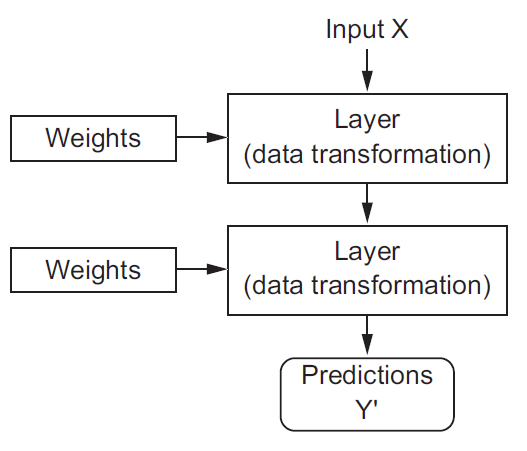


Weights Layer (data transformation) Layer (data transformation)

Predictions Y’ Weights

1. Explain how “transformation implemented by a layer is parameterized by its weights”.
2. What computes the error of the prediction to the ‘true target’? Why is this important? What is this iteration called?
3. Deep learning is the best tool to solve most if not all problems. (True or False). Why?
4. How does a support vector machine (SVM) solve classification problems?
5. What are the difficulties in using SVMs?
   1. hard to scale
   2. do not provide good results for perceptual problems
   3. shallow method
   4. uses feature engineering
   5. All the above
6. What is a kernel function?
7. What does deep learning automate that is a crucial step in a machine-learning workflow?
8. What is joint feature learning?
9. What are the two essential characteristics of how deep learning learns from data?
10. What are the 3 technical forces driving advances in machine learning?
11. How are GPU developments and deep learning related?
12. What is a TPU?
13. What are the categories of properties that make deep learning a viable long-term option?

Answers 1

1. –
   1. Classical programming
   2. Machine learning
2. False
3. C. Alan Turing 1950
4. A. large and complex datasets.
5. C, A, B
6. A
7. B, An automatic search process for better representations
8. A, A predefined set of hypotheses
9. C, Layered representations in deep learning
10. Shallow learning tends to focus on 1 to 2 layers of representations while deep learning can involve tens or hundreds.
11. Deep learning is a multistage information-distillation operation. There are success filters and the information comes out increasingly purified.
12. 
13. As data is transformed through each layer in the deep learning process, the weight that each layer has affects how much the data is transformed.
14. The loss function. Measuring the distance between the output and expectation allows the weights to be adjusted accordingly with the goal of decreasing the loss score. Training loop.
15. (15) False. Sometimes there isn’t enough data. Other algorithms could fit the problem better.
16. (15) SVMs use decision boundaries that maximize the margin (the distance between the hyperplane and the closest data points from each class).
17. (p. 16) E, all the above
18. (16) “A kernel function is a computationally tractable operation that maps any two points in your initial space to the distance between these points in your target representation space, completely bypassing the explicit computation of the new representation. Kernel functions are typically crafted by hand rather than learned from data—in the case of an SVM, only the separation hyperplane is learned.”
19. (18) Feature engineering
20. (18) When a model learns(adjusts) all layers of representation together. “The incremental, layer-by-layer way in which increasingly complex representations are developed
21. (18) “the *incremental, layer-by-layer way in which increasingly complex representations are developed*, and the fact that *these intermediate incremental representations are learned jointly*”
22. (20) hardware, datasets and benchmarks, algorithmic advances
23. (20) Both are highly parallelizable.
24. (21) Google’s tensor processing unit (TPU), “a new chip design developed from the ground up to run deep neural networks, which is reportedly 10 times faster and far more energy efficient than top-of-the-line GPUs.”
25. (23) Simplicity, scalability, versatility and reusability