1. Do you need these for deep learning?
   * Lots of math T / **F**
   * Lots of data T / **F**
   * Lots of expensive computers T / **F**
   * A PhD T / **F**
2. Name five areas where deep learning is now the best in the world.  
   NLP, Computer vision, medicine, biology, image generation, recommendation systems, playing games, robotics.
3. What was the name of the first device that was based on the principle of the artificial neuron?  
   Mark I Perceptron.
4. Based on the book of the same name, what are the requirements for parallel distributed processing (PDP)?   
   - A set of processing units,   
   - a state of activation,   
   - an output function for each unit,   
   - a pattern of connectivity among units,   
   - a propagation rule for propagating pattern of activities through the network of connectivities,   
   - an activation rule for combining the inputs impinging on a unit with the current state of that unit to produce an output for the unit,  
   - a learning rule whereby patterns of connectivity are modified by experience,  
   - an environment within which the system must operate.
5. What were the two theoretical misunderstandings that held back the field of neural networks?  
   - In 1969, neural networks are thought to unable to learn some simple but critical mathematical functions (such as XOR) – in truth, it was due to only using a single layer instead of multiple layers on perceptron.  
   - In 1980, networks are too big and too slow to approximate any mathematical functions using two layers of artificial neurons. While using additional layers improved performance, this insight was not acknowledged.
6. What is a GPU?   
   Also known as a graphics card, a special kind of processor in computer that can handle thousands of single tasks at the same time, especially designed for displaying 3D environments on a computer for playing games. These same basic tasks are very similar to what neural networks do, such that GPUs can run neural networks hundreds of times faster than regular CPUs. All modern computers contain a GPU, but few contain the right kind of GPU necessary for deep learning.
7. Open a notebook and execute a cell containing: 1+1. What happens?   
   Jupyter will show the result (2).
8. Follow through each cell of the stripped version of the notebook for this chapter. Before executing each cell, guess what will happen.  
   Depends on the code.
9. Complete the Jupyter Notebook online appendix.  
   Testing models with hotels and restaurants.
10. Why is it hard to use a traditional computer program to recognize images in a photo?  
    Because traditional computer program needs to be programmed exactly step-by-step. However, it is really hard to program such steps for recognizing images, just as humans do not recognize images step-by-step, but automatically by our brains.
11. What did Samuel mean by "weight assignment"?  
    Weights are just variables, and a weight assignment is a particular choice of values for those variables. The program’s inputs are values that it processes in order to produce its results – for instance, taking image pixels as inputs and returning the classification “dog” as a result. The program’s weight assignments are other values that define how the program will operate.
12. What term do we normally use in deep learning for what Samuel called "weights"?  
    Model *parameters*.
13. Draw a picture that summarizes Samuel's view of a machine learning model.  
    Inputs + weights -> model -> results -> performance –update🡪 weights.
14. Why is it hard to understand why a deep learning model makes a particular prediction?  
    Because the mechanism for altering the weight assignment to maximize the performance is automatic – the machine would learn from its experience instead of being adjusted by us. Because of the “deep” nature, layers can have hundreds or thousands of layers. It’s hard to determine which factors are important in determining the final output.
15. What is the name of the theorem that shows that a neural network can solve any mathematical problem to any level of accuracy?  
    Universal approximation theorem.
16. What do you need in order to train a model?  
    Inputs (training data, validation data, test sets – each data having labels), architecture, parameters, loss function, a way to update the parameters of the model to improve its performance (optimizer).
17. How could a feedback loop impact the rollout of a predictive policing model?  
    - This is the case when the more the model is used, the more biased the data becomes, making the model even more biased, and so forth.   
    - Example is when trying to predict crimes, but in actuality is predicting arrests. Law enforcement officers might use that model to decide where to focus their police activity, resulting in increased arrests in those areas, and data on these additional arrests would then be fed back in to retrain future versions of the model.
18. Do we always have to use 224×224-pixel images with the cat recognition model?  
    No. It is only used because it is the standard size for historical reasons (old pretrained models require this size exactly), but you can pass pretty much anything. Increasing the size will have better results (since it will be able to focus on more details), but at the price of speed and memory consumption; the opposite is true if size is decreased.
19. What is the difference between classification and regression?  
    - A classification model is one which attempts to predict a class or category, predicting from a number of discrete possibilities, such as “dog” or “cat”.  
    - A regression model is one which attempts to predict one or more numeric quantities, such as a temperature or a location.
20. What is a validation set? What is a test set? Why do we need them?  
    - Validation set is used outside training set and is used to evaluate whether our model makes a correct prediction to data that the model has never seen.  
    - Test set is used to prevent us humans to have our bias when improving our model. This test set is a set that we should not see at all, only used to evaluate the model at the very end of our efforts.  
    - We use validation and test set to prevent bias of both machine and humans when training and improving the model.
21. What will fastai do if you don't provide a validation set?  
    It requires validation set to run correctly, so fastai will randomly take 20% of the training data to become validation set.
22. Can we always use a random sample for a validation set? Why or why not?  
    Depends on the case, it may be not a good idea to use a random sample in a time series data (as it will not be indicative of what you’ll see in production, which is the future data), or in the case where the production’s predictions may be qualitatively different from the data you have to train your model with (such as in distracted driver competition, you need to make sure that the model is not overfit with a small group of people, so the people in the validation set needs to be different from the people in the training set).
23. What is overfitting? Provide an example.  
    When the model has a very good accuracy in training set, but not the validation set, as the model starts to memorize the training set, rather than finding a generalizable underlying patterns in the data. Example is during Kaagle fisheries competition, it may be possible to overfit to the boats’ images in the training set.
24. What is a metric? How does it differ from "loss"?  
    - A metric is a function that measures the quality of the model’s predictions using the validation set, and will be printed at the end of each “epoch”. In this case, we’re using “error\_rate”, which is a function provided by fastai that does just what it says: tells you what percentage of images in the validation set are being classified incorrectly.  
    - The entire purpose of loss is to define a “measure of performance” that the training system can use to update weights automatically. In other words, a good choice for loss is a choice that is easy for stochastic gradient descent to use. But a metric is defined for human consumption, so a good metric is one that is easy for you to understand, and that hews as closely as possible to what you want the model to do. At times, you might decide that the loss function is a suitable metric, but that is not necessarily the case.
25. How can pretrained models help?  
    - As pretrained models have weights that have already been trained on some other dataset, they are very capable and will handle edge, gradient, and color detection, which are needed for many tasks. The allow us to train more accurate models, more quickly, with less data, and less time and money. Using pretrained model for a task different to what it was originally trained for is known as transfer learning.
26. What is the "head" of a model?  
    Last part (or layers) of the model, which is specific to the (new) dataset.
27. What kinds of features do the early layers of a CNN find? How about the later layers?  
    - Early layers: The model has discovered weights that represent diagonal, horizontal, and vertical edges, as well as various different gradients. Next layer, the model has learned to create feature detectors that look for corners, repeating lines, circles, and other simple patterns.  
    - Later layers: The features are now able to identify and match with higher-level semantic components, such as car wheels, text, and flower petals. Next layer can identify even higher-level concepts.
28. Are image models only useful for photos?  
    No, for example it is possible for a sound to be converted to a spectrogram, which is a chart that shows the amount of each frequency at each time in an audio file. The spectrogram can be used as a dataset for environmental sound detection model. Another example is detecting malware by changing the 8-bit sequences (malware binary) to equivalent decimal values, which are then reshaped and a gray-scale image is generated that represents the malware sample.
29. What is an "architecture"?   
    The template of the model that we’re trying to fit; the actual mathematical function that we’re passing the input data and parameters to.
30. What is segmentation?  
    Creating a model that can recognize the content of every individual pixel in an image is called segmentation. For example, it is critically important for autonomous vehicles to localizing objects in a picture.
31. What is y\_range used for? When do we need it?  
    When we’re predicting a continuous number, rather than a category, we have to tell fastai what range our target has, using the y\_range parameter.
32. What are "hyperparameters"?  
    Parameters about parameters, the context is we are likely to explore many versions of a model through various modeling choices regarding network architecture, learning rates, data augmentation strategies, and other factors. Many of these choices can be described as choices of hyperparameters, since they are the higher-level choices that govern the meaning of the weight parameters.
33. What's the best way to avoid failures when using AI in an organization?  
    - Ensure that you really understand what test and validation sets are and why they’re important. If you’re considering bringing in an external vendor or service, make sure that you hold out some test data that the vendor never gets to see. Then you check their model on your test data using a metric that you choose based on what actually matters to you in practice, and you decide what level performance is adequate.  
    - Try out a simple baseline, which future models should hopefully beat. Or even this simple baseline may be enough in some cases.

Further Research

1. Why is a GPU useful for deep learning? How is a CPU different, and why is it less effective for deep learning?  
   - GPU is designed to parallelization capabilities; execute millions of operations in parallel (while sacrificing speed of individual threads). This capability is crucial for efficiently performing matrix multiplications and other fundamental operations in deep learning algorithms.  
   - CPU is designed to execute a sequence of operations (thread) as fast as possible (and can only execute dozens of them simultaneously).
2. Try to think of three areas where feedback loops might impact the use of machine learning. See if you can find documented examples of that happening in practice.  
   - People finding what’s popular by seeing what are people using. Train the data, so more people will follow what’s popular, it gets more popular, another retraining, and the data get even more biased.  
   - If black men in Baltimore are more likely to go to jail because of the color of their skin and their sex, then the model will pick this up.  
   - If we add gender to our hiring models, and the vast majority of “successful” software developers happen to be male, then should we only pick the male candidates?

Answer: <https://forums.fast.ai/t/fastbook-chapter-1-questionnaire-solutions-wiki/65647>

Reference: <https://towardsdatascience.com/dangerous-feedback-loops-in-ml-e9394f2e8f43>