

## **FastAI Course Lesson 1**

**Video Series:** 1/9

**Book:** Chp 1/20

# Summary Qs

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

### **1. Name five areas where deep learning is now the best in the world**

- Natural Language Processing (NLP) – Question Answering, Document Summarization and Classification, etc.
- Computer Vision – Satellite and drone imagery interpretation, face detection and recognition, image captioning, etc.
- Medicine – Finding anomalies in medical images (ex: CT, X-ray, MRI), detecting features in tissue slides (pathology), diagnosing diabetic retinopathy, etc.
- Biology – Folding proteins, classifying, genomics tasks, cell classification, etc.
- Image generation/enhancement – colorizing images, improving image resolution (super-resolution), removing noise from images (denoising), converting images to art in the style of famous artists (style transfer), etc.
- Recommendation systems – web search, product recommendations, etc.
- Playing games – Super-human performance in Chess, Go, Atari games, etc
- Robotics – handling objects that are challenging to locate (e.g. transparent, shiny, lack of texture) or hard to pick up
- Other applications – financial and logistical forecasting; text to speech; much much more.

### **2. What was the name of the first device that was based on the principle of the artificial neuron?**

Mark I perceptron built by Frank Rosenblatt

### **3. Based on the book of the same name, what are the requirements for “Parallel Distributed Processing”?**

- 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 patterns 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 a new level of activation for the unit
- A learning rule whereby patterns of connectivity are modified by experience
- An environment within which the system must operate

### **4. What were the two theoretical misunderstandings that held back the field of neural networks?**

In 1969, Marvin Minsky and Seymour Papert demonstrated in their book, “Perceptrons”, that a single layer of artificial neurons cannot learn simple, critical mathematical functions like XOR logic gates. While they subsequently demonstrated in the same book that additional layers can solve this problem, only the first insight was recognized, leading to the start of the first AI winter.

In the 1980’s, models with two layers were being explored. Theoretically, it is possible to approximate any mathematical function using two layers of artificial neurons. However, in practice, these networks were too big and too slow. While it was demonstrated that adding additional layers improved performance, this insight was not acknowledged, and the second AI winter began. In this past decade, with increased data availability, and improvements in computer hardware (both in CPU performance but more importantly in GPU performance), neural networks are finally living up to their potential.

## **5. What is a GPU?**

CPUs, or central processing units, are the core units of all standard computers, and they execute the instructions that make up computer programs. A Graphics Processing Unit (also known as a graphics card) are specialized units meant for displaying graphics, especially the 3D graphics in modern computer games. The hardware optimizations used in GPUs allow it to handle thousands of tasks at the same time. Incidentally, these optimizations allow us to run and train neural networks hundreds of times faster than a regular CPU.

## **6. Why is it hard to use a traditional computer program to recognize images in a photo?**

For us humans, it is easy to identify images in photos, such as identifying cats vs dogs in a photo. This is because, subconsciously our brains have learned which features define a cat or a dog for example. But it is hard to define set rules for a traditional computer program to recognize a cat, dog or other objects since they have a wide variety of shapes, textures, colors, and other features. It's close to impossible to manually encode this in a traditional computer program.

## **7. What did Samuel mean by “Weight Assignment”?**

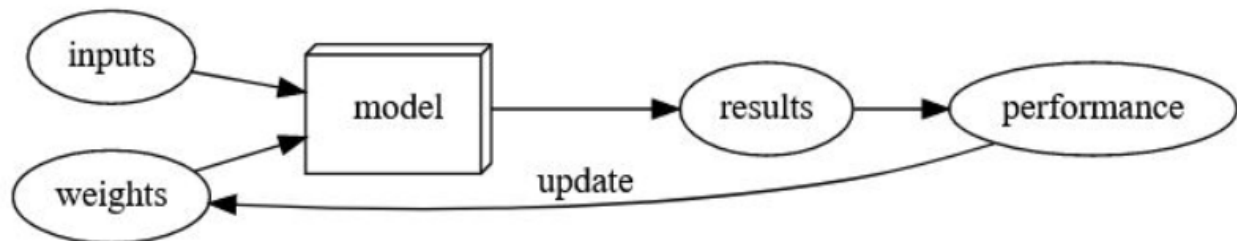
Refers to the current values of the model parameters. Arthur Samuel further mentions an *“automatic means of testing the effectiveness of any current weight assignment”* and a *“mechanism for altering the weight assignment so as to maximize the performance”*. This refers to the evaluation and training of the model in order to obtain a set of parameter values that maximizes model performance.

## **8. What term do we normally use in deep learning for what Samuel called “Weights”?**

We instead use the term parameters. In deep learning, the term “weights” has a separate meaning. (The neural network has various

parameters that we fit our data to. As shown in upcoming chapters, the two types of neural network parameters are weights and biases)

**9. Draw a picture that summarizes Arthur Samuel's view of a machine learning model**



**10. Why is it hard to understand why a deep learning model makes a particular prediction?**

This is a highly-researched topic known as interpretability of deep learning models. Deep learning models are hard to understand in part due to their “deep” nature. Think of a linear regression model. Simply, we have some input variables/data that are multiplied by some weights, giving us an output. We can understand which variables are more important and which are less important based on their weights. A similar logic might apply for a small neural network with 1-3 layers. However, deep neural networks have hundreds, if not thousands, of layers. It is hard to determine which factors are important in determining the final output. The neurons in the network interact with each other, with the outputs of some neurons feeding into other neurons. Altogether, due to the complex nature of deep learning models, it is very difficult to understand why a neural network makes a given prediction. However, in some cases, recent research has made it easier to better understand a neural network’s prediction. For example, as shown in this chapter, we can analyze the sets of weights and determine what kind of features activate the neurons. When applying CNNs to images, we can also see which parts of the images highly activate the model. We will see how we can make our models interpretable later in the book.

**11. What is the name of the theorem that a neural network can solve any mathematical problem to any level of accuracy?**

The universal approximation theorem states that neural networks can theoretically represent any mathematical function. However, it is important to realize that practically, due to the limits of available data and computer hardware, it is impossible to practically train a model to do so.

**12. What do you need in order to train a model?**

- “Architecture” for the given problem (Architecture is the template or structure of the model we are trying to fit. It defines the mathematical model we are trying to fit)
- Data to input to your model
- Labels for your data to compare your model predictions to
- Loss function that will quantitatively measure the performance of your model
- Way to update the parameters of the model in order to improve its performance (i.e. optimizer).

**13. How could a feedback loop impact the rollout of a predictive policing model?**

In a predictive policing model, we might end up with a positive feedback loop, leading to a highly biased model with little predictive power. For example, we may want a model that would predict crimes, but we use information on arrests as a *proxy*. However, this data itself is slightly biased due to the biases in existing policing processes. Training with this data leads to a biased model. Law enforcement might use the model to determine where to focus police activity, increasing arrests in those areas. These additional arrests would be used in training future iterations of models, leading to an even more biased model. This cycle continues as a positive feedback loop

**14. Do we always have to use 224x224 pixel images with the cat recognition model?**

224x224 is commonly used for historical reasons. You can increase the size and get better performance, but at the price of speed and memory consumption.

**15. What is the difference between classification and regression?**

Classification = predicting a class or category (ex: type of pet)

Regression = predicting a numeric quantity (ex: age of pet)

**16. What is a validation set? What is a test set? Why do we need them?**

The validation set is the portion of the dataset that is not used for training the model, but for evaluating the model during training, in order to prevent overfitting. This ensures that the model performance is not due to “cheating” or memorization of the dataset, but rather because it learns the appropriate features to use for prediction. Fastai randomly selects 20% of the data and assigns it as the validation set (`valid_pct = 0.2`).

However, it is possible that we overfit the validation data as well. This is because the human modeler is also part of the training process, adjusting “hyperparameters” and training procedures according to the validation performance. Therefore, another unseen portion of the dataset, the test set, is used for final evaluation of the model. This splitting of the dataset is necessary to ensure that the model generalizes to unseen data.

**17. What are “hyperparameters”?**

Training models require various other parameters that define how the model is trained. For example, we need to define how long we train for, or what learning rate (how fast the model parameters are allowed to change) is used.

**18. What is overfitting? Provide an example.**

Overfitting refers to when the model fits too closely to a limited set of data but does not generalize well to unseen data. This is especially important when it comes to neural networks, because neural networks can potentially “memorize” the dataset that the model was trained on, and will perform abysmally on unseen data because it didn’t “memorize” the ground truth values for that data. This is why a proper validation framework is needed by splitting the data into training, validation, and test sets.

**19. Can we always use a random sample for a validation set? Why or why not?**

A good validation or test set should be representative of new data you will see in the future. Sometimes this isn’t true if a random sample is used. For example, for time series data, selecting sets randomly does not make sense. Instead, defining different time periods for the train, validation, and test set is a better approach.

**20. 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. This is similar to the loss, which is also a measure of performance of the model. However, loss is meant for the optimization algorithm (like SGD) to efficiently update the model parameters, while metrics are human-interpretable measures of performance. Sometimes, a metric may also be a good choice for the loss.

**21. How can pre-trained models help?**

Pre-trained models have been trained on other problems that may be quite similar to the current task. For example, pretrained image recognition models were often trained on the ImageNet dataset, which has 1000 classes focused on a lot of different types of visual

objects. Pretrained models are useful because they have already learned how to handle a lot of simple features like edge and color detection. However, since the model was trained for a different task than already used, this model cannot be used as is.

## **22. What is the “head” of a model?**

When using a pretrained model, the later layers of the model, which were useful for the task that the model was originally trained on, are replaced with one or more new layers with randomized weights, of an appropriate size for the dataset you are working with. These new layers are called the “head” of the model.

## **23. What kinds of features do the early layers of a CNN find? How about the later layers?**

Earlier layers learn simple features like diagonal, horizontal, and vertical edges. Later layers learn more advanced features like car wheels, flower petals, and even outlines of animals.

## **24. Are image models only useful for photos?**

Nope! Image models can be useful for other types of images like sketches, medical data, etc. However, a lot of information can be represented as images. For example, a sound can be converted into a spectrogram, which is a visual interpretation of the audio. Time series (ex: financial data) can be converted to image by plotting on a graph. Even better, there are various transformations that generate images from time series, and have achieved good results for time series classification. There are many other examples, and by being creative, it may be possible to formulate your problem as an image classification problem, and use pretrained image models to obtain state-of-the-art results!

## **25. What is segmentation?**



At its core, segmentation is a pixel wise classification problem. We attempt to predict a label for every single pixel in the image. This provides a mask for which parts of the image correspond to the given label.

**26. What is y\_range used for? When do we need it?**

y\_range is being used to limit the values predicted when our problem is focused on predicting a numeric value in a given range (ex: predicting movie ratings, range of 0.5-5).

**27. What's the best way to avoid failures when using AI in an organization?**

- Make sure a training, validation, and testing set is defined properly in order to evaluate the model in an appropriate manner.
- Try out a simple baseline, which future models should hopefully beat. Or even this simple baseline may be enough in some cases.