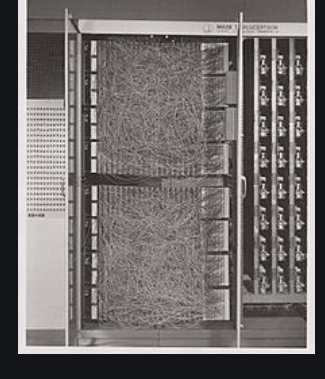
1. Do you need these for deep learning?
   * Lots of math T / F : False
   * Lots of data T / F : False
   * Lots of expensive computers T / F : False
   * A PhD T / F False
2. Name five areas where deep learning is now the best in the world.
   * + Natural language processing
     + Computer vsion
     + Medicine
     + Biology
     + Image generation
     + Recommendation System
     + Robotics
3. What was the name of the first device that was based on the principle of the artificial neuron? Mark I Perceptron



1. 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 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 an output for the unit
* A *learning rule* whereby patterns of connectivity are modified by experience
* An *environment* within which the system must operate

1. What were the two theoretical misunderstandings that held back the field of neural networks?
   * + They showed that a single layer of these devices was unable to learn some simple but critical mathematical functions (such as XOR)
     + In theory, adding just one extra layer of neurons was enough to allow any mathematical function to be approximated with these neural networks, but in practice such networks were often too big and too slow to be useful.
2. What is a GPU?
   * Graphic Processing Unit, muh faster forthan unit in term of neural network
3. Open a notebook and execute a cell containing: 1+1. What happens?

It prints out 2

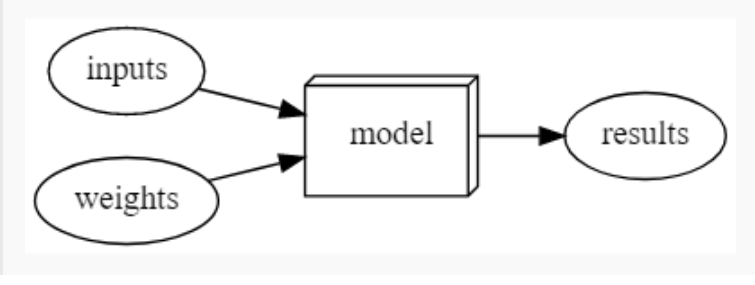
1. Follow through each cell of the stripped version of the notebook for this chapter. Before executing each cell, guess what will happen.
2. Complete the Jupyter Notebook online appendix.
3. Why is it hard to use a traditional computer program to recognize images in a photo?
   * + Universal rules is very diffucult to recognize images
4. What did Samuel mean by "weight assignment"?

“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.

1. What term do we normally use in deep learning for what Samuel called "weights"?

parameters

1. Draw a picture that summarizes Samuel's view of a machine learning model.



1. Why is it hard to understand why a deep learning model makes a particular prediction?
   * There are very many many layer. It is easy if there is few layer, but many later it is different thing
2. What is the name of the theorem that shows 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. But we can get very close!

1. What do you need in order to train a model?
   * Architecture(function)
   * Input
   * Label for prediction
   * Loss function
   * way to update the parameters of the model in order to improve its performance (this is known as an optimizer).
2. How could a feedback loop impact the rollout of a predictive policing model?
   * It will continuous affect the decision of the poliec on that location and further loop 🡪 the same location
3. 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*
4. Do we always have to use 224×224-pixel images with the cat recognition model?
   * It is universal for the old time but we don’t have to use 224 every time
5. What is the difference between classification and regression?
   * Regression : number
   * Classifiation : category
6. What is a validation set? What is a test set? Why do we need them?
   * Valdiate is to use after we train the model
   * Test set is used after we validate, it is hidden entirely
   * We use test set to reduce the validation bias

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. 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 (see question 32 for definition) 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.

1. What will fastai do if you don't provide a validation set?

It will automatially do 20% ccut validation data set

1. Can we always use a random sample for a validation set? Why or why not?
   * Not really, sometimes it is the same data with the same pattern
   * 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.
2. What is overfitting? Provide an example.
   * Overfitting is the most challenging issue when it comes to training machine learning models. 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.
3. What is a metric? How does it differ from "loss"?
   * A metric is a function that measures 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.
4. How can pretrained models help?
   * Pretrained 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.
5. 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.
6. 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.
7. Are image models only useful for photos?
   * No, you can do it in security as well and any other applications
   * 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!
8. What is an "architecture"?
   * Function
   * The architecture is the template or structure of the model we are trying to fit. It defines the mathematical model we are trying to fit.
9. What is segmentation?
   * At its core, segmentation is a pixelwise 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.
10. What is y\_range used for? When do we need it?
    * To limit the value of the dependent value
    * 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).
11. What are "hyperparameters"?

Training models requires 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. These sorts of parameters are hyperparameters.

1. 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.( VERY IMPORTANT)
* Try out a simple baseline, which future models should hopefully beat. Or even this simple baseline may be enough in some cases.( THE GOAL)

1. Why is a GPU useful for deep learning? How is a CPU different, and why is

* Gpu is useful for computation tasks on graphic, which is perfect because deep learning using a lot of picture and images to train.
* GPU is generally faster than CPU

1. Try to  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.
   1. In business field: Stock market can be affected by machine learning, leading to the same result and incline to that trend when automatic system buying stock
   2. Decision in business, buying good deal from this company and never buy other good deal from other company
   3. Insurance: only for people with 50K NW or white