# Lesson 1 '01\_intro.ipynb'

## 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

You need none of these.

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

I. Robotics - handling objects that are challenging to locate (e.g. transparent, shiny, lack of texture) or hard to pick up

II. Computer vision - satellite and drone imagery interpretation (e.g. for disaster resilience); face recognition; image captioning; reading traffic signs; locating pedestrians and vehicles in autonomous vehicles

III. Image generation - Colorizing images; increasing image resolution; removing noise from images; converting images to art in the style of famous artists

IV. Natural Language Processing (NLP) - answering questions; speech recognition; summarizing documents; classifying documents; finding names, dates, etc. in documents; searching for articles mentioning a concept

V. Other applications - financial and logistical forecasting; text to speech; much much more..

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

Perceptron

## 4. Based on the book of the same name, what are the requirements for "Parallel Distributed Processing"?

1. A set of processing units
2. A state of activation
3. An output function for each unit
4. A pattern of connectivity among units
5. A propagation rule for propagating patterns of activities through the network of connectivities
6. 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
7. A learning rule whereby patterns of connectivity are modified by experience
8. An environment within which the system must operate

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

1. Adding just one extra layer of neurons was enough to allow any mathematical function to be approximated with these neural networks.
2. Although researchers showed 30 years ago that to get practical good performance you need to use even more layers of neurons, it is only in the last decade that this principle has been more widely appreciated and applied.

## 6. What is a GPU?

GPU, also known as a *graphics card*, is aspecial kind of processor in your computer than 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, following through each cell of the stripped version of the notebook for this chapter. Before executing each cell, guess what will happen.

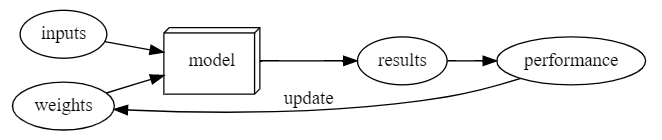
## What did Samuel mean by "Weight Assignment"?

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. But the program's weight assignments are other values which define how the program will operate.

## What term do we normally use in deep learning for what Samuel called "Weights"?

Model parameters

## Draw a picture that summarizes Arthur Samuel's view of a machine learning model



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

Universal approximation theorem

## What do you need in order to train a model?

Dataset and a label for each data

## How could a feedback loop impact the rollout of a predictive policing model?

The feedback loop makes the model biased, it is not predicting the outcome of data but the outcome based on previous results thus reflecting biases in the existing policy. The more biased the data becomes the more biased it makes the model and so forth.

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

No, you can pass pretty much anything. If you increase the size, you'll often get a model with 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 you decrease the size.

## What is the difference between classification and regression?

A classification model is one which attempts to predict a class, or category. That is, it's 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.

## What is a validation set? What is a test set? Why do we need them?

The validation set is used to measure the accuracy of the model. By default, say 20%of the input data is selected randomly to test the model accuracy.

The test set is used to evaluate the model at the very end of the modeling process. If we define a hierarchy of cuts of our data, the training data is fully exposed, the validation data is less exposed, and test data is totally hidden.

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

Fastai will choose a validation set of 20% using parameter ‘valid\_pct = 0.2’ and set the parameter ‘seed’ to 42 so that the same validation set is chosen every time we run the model.

## Can we always use a random sample for a validation set? Why or why not?

No, we need to fix the validation set. This will help us test the model performance. If we change our model, we know that any differences are due to the changes to the model, not due to having a different random validation set.

## What is overfitting? Provide an example.

If you train your model longer it will start to memorize the training set, rather than finding generalizable underlying patterns in the data. When this happens, we say that the model is *overfitting*.

A model architecture called *ResNet* is both fast and accurate for many datasets and problems. The 34 in resnet34 refers to the number of layers in this variant of the architecture (other options are 18, 50, 101, and 152). Models using architectures with more layers take longer to train, and are more prone to overfitting

## What is a metric? How does it differ to "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* (one complete pass through the dataset).

The purpose of loss is to define a "measure of performance" that the training system can use to update weights automatically. 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.

## How can pretrained models help?

Pretrained models allow us to train more accurate models, more quickly, with less data, and less time and money.

Using a pretrained model for a task different to what it was originally trained for is known as *transfer learning*.

## What is the "head" of a model?

The head of a model is the part that is newly added to be specific to the new dataset.

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

For layer 1, what we can see is that the model can discover weights that represent diagonal, horizontal, and vertical edges, as well as various different gradients.

For layer 2, the model can create feature detectors that look for corners, repeating lines, circles, and other simple patterns.

For layer 3, the features identify and match with higher-level semantic components, such as car wheels, text, and flower petals. Using these components, layers four and five can identify even higher-level concepts.

## Are image models only useful for photos?

No, an image model can learn to complete many tasks. For example

1. A sound can be converted to a spectrogram, which is a chart that shows the amount of each frequency at each time in an audio file.
2. A time series can easily be converted into an image by simply plotting the time series on a graph.

## What is an "architecture"?

The functional form of the *model* is called its architecture. The architecture only describes a *template* for a mathematical function; it doesn't actually do anything until we provide values for the millions of parameters it contains.

## What is segmentation?

Creating a model that can recognize the content of every individual pixel in an image is called segmentation.

## What is y\_range used for? When do we need it?

y\_range is used in a regression dataset where we are predicting a continuous number, rather than a category.

## What are "hyperparameters"?

In realistic scenarios we rarely build a model just by training its weight parameters once. Instead, we explore many versions of a model through various modeling choices regarding network architecture, learning rates, data augmentation strategies etc. Many of these choices can be described as choices of *hyperparameters*. The word reflects that they are parameters about parameters, since they are the higher-level choices that govern the meaning of the weight parameters

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

You ensure that you really understand what test and validation sets are and why they're important. For instance, 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 of performance is adequate.

## Try to think of three areas where feedback loops might impact use of machine learning. See if you can find documented examples of that happening in practice.

I will try to illustrate positive implications of feedback loops in machine learning here

1. Problems in healthcare where there is relatively rich and continuous collection of unbiased data, and a tight feedback loop between intervention and response. One example is the adjustment of diabetic medications to reduce diabetic medications by measuring multiple glucose readings per day. In each case, a drug/medical device is adjusted in near real-time based on biomarker based feedback that completes a feedback loop.
2. Active Learning is a special case of machine learning in which a learning algorithm uses the feedback loop for labeling the data. In Active learning you start with an unlabeled dataset and split data into a very small dataset which gets a label and a large unlabeled dataset. In active learning terminology that small labelled dataset is called as seed, next train initial model on the seed dataset and predict the labels of the remaining unlabeled observations. Lastly use the uncertainty of the model's predictions to prioritize the labeling of remaining observations. [Active learning (machine learning)](https://en.wikipedia.org/wiki/Active_learning_(machine_learning)) and [[2002.05033] Active Learning for Sound Event Detection](https://arxiv.org/abs/2002.05033)
3. Robotic systems with intelligent automation systems can control complex tasks in unstructured environments such as surface mining or construction. [Incorporating Expert Feedback into Active Anomaly Discovery - IEEE Conference Publication](https://ieeexplore.ieee.org/document/7837915) and [Electrical Engineering and Systems Science authors/titles recent submissions](https://arxiv.org/list/eess/recent)