Answer for question 1:

The major difference between classical machine learning and representation learning is that for classical machine learning, we define the features for it, and it starts learning based on these features that provided by humans, but for representation learning, it learns its features by itself.

And learned representations often results in much better performance than what can be obtained with hand-designed representations.

The power of good representation can be seen by looking at the next figure.

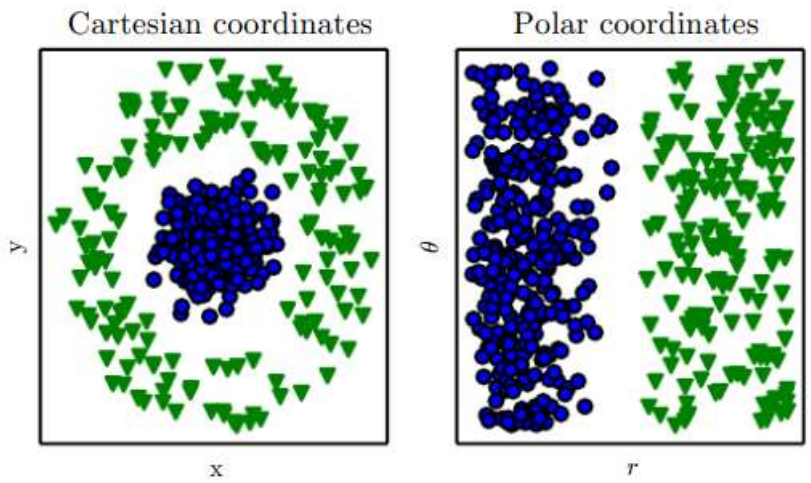

Figure 1: Example of different representations

As the data can't be separated using the cartesian coordinates by a line but if we used another representation, polar coordinates, we can.

And to separate the factors of variation with designing algorithms for learning features won't usually end up with the same features that is extracted by humans as not all these features are easily observed but it may be unobserved forces in the physical world that affect physical quantities.

Answer for question 2:

It's a truth that some skill is required to get good performance from a deep learning algorithm, but the amount of skill required reduces as the amount of training data increases.

A great proof that the size of datasets mainly affects the efficiency of the results is that the learning algorithms used nowadays and reaching human performance are almost the same learning algorithms that struggled to solve simpler problems back in 1980s.

And regarding the datasets that has contributed to the evolution of deep learning, the following figure shows these datasets along time.

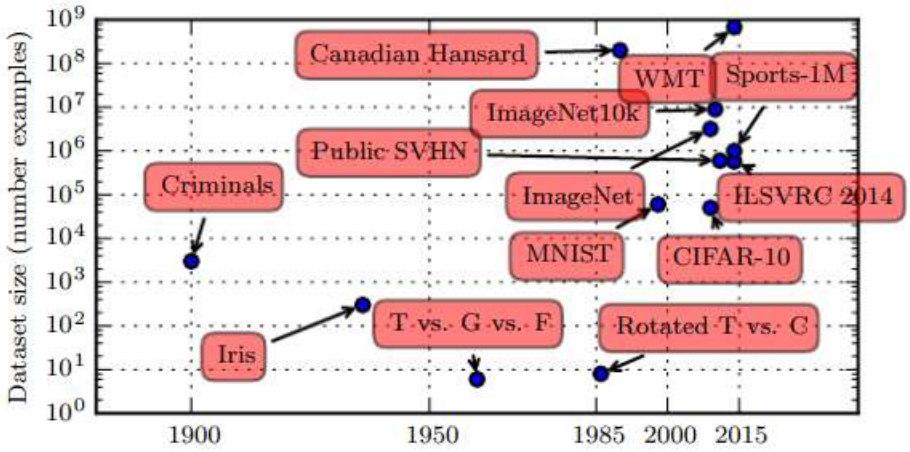

Figure 2: Increasing datasets over time.

- In early 1900s, statisticians studied datasets using hundreds or thousands of manually compiled measurements.
- In the 1950s through the 1980s, the pioneers of biologically inspired machine learning often worked with small synthetic datasets, such as low-resolution bitmaps of letters, that were designed to incur low computational cost and demonstrate that neural networks were able to learn specific kinds of functions.
- In the 1980s and 1990s, machine learning became more statistical and began to leverage larger datasets containing tens of thousands of examples, such as the **MNIST dataset**.
- In the first decade of the 2000s, more sophisticated datasets of this same size, such as the **CIFAR-10** dataset continued to be produced.
- Toward the end of that decade and throughout the first half of the 2010s, significantly larger datasets, containing hundreds of thousands to tens of millions of examples, completely changed what was possible with deep learning. These datasets included:
    - The **public Street View House Numbers dataset**.
    - Various versions of the **ImageNet** dataset.
    - The **Sports-1M** dataset.

Dataset	MNIST	**CIFAR-10**	**public Street View House Numbers (public SVHN) dataset**	**ImageNet**	**Sports-1M**
Size	60,000 grayscale images	60000 32x32 colour images	Over 600,000-digit images	1,133,158 video URLs	1.2 million images

Answer for question 3

2 factors mainly control the capacity of neural network:

- Number of nodes.
- Number of layers.

A model with more layers and more hidden units per layer has higher representational capacity — it is capable of representing more complicated functions.

Answer for question 4

$$\arg_\theta \min \frac{1}{T} \sum_T l\left(X^{(t)}; \theta\right), y^{(t)}) + \lambda \Omega(\theta)$$

T: Number of training example.

$l\left(X^{(t)}; \theta\right), y^{(t)})$: Loss function

$\Omega(\theta)$: Regulaizer.

The idea in machine learning is to learn a function that minimizes the error or one that maximizes the benefit and Finding the parameters for machine learning can be thought of as an optimization for the model but with more optimization we can find our model overfitting to the training set which leads to unwanted results but a normal optimization problem normally, with more optimization, lead to better results.

Also, for optimization problems casually we need mathematical

Answer for question 5

https://colab.research.google.com/drive/1lfuwbpDLf92lckJmxlOxWw9miTx5LQpn?usp=sharing

Answer for question 6

https://colab.research.google.com/drive/1V41CYTDIKomvqxzf165GS99sygrUdNbp?usp=sharing

Reference:

- Lecture slides.
- Deep Learning (deeplearningbook.org)
- A Few Words on Representation Learning - Thalles' blog (sthalles.github.io)
- How to Control Neural Network Model Capacity With Nodes and Layers (machinelearningmastery.com)