Introduction

Early on

Early on, artificial intelligence tackled problems that were easily formalized, though difficult or unintuitive for humans. However, things that are intuitive for humans but difficult to formalize (eg. Recognizing spoken word, faces in images) posed challenges.

The solution was to allow computers to learn from experience, understand the world in terms of a hierarchy of concepts. The ultimate goal was to get informal knowledge into a computer. Because a graph of concept relations would have many layers, the term “deep” learning was coined.

Approaches

The *knowledge base approach* to artificial intelligence tries to hard-code knowledge into formal statements (eg. Cyc project (1989) attempted). These projects were not very successful because they still required humans to do the formalizing.

Alternatively, *machine learning* involves machines examining a representation of data. Several pieces of information (features) about a entity are included in the record that the model is run on. Thus, the basic constraint is that the representation must be chosen.

A solution to this is to learn the representation itself (*representation learning*). *Autoencoders* are an example of a representation learning algorithm. Autoencoders learn features by converting data to some representation using an *encoder function* and back with a *decoder function*, trying to preserve as much information as possible.

Factors of Variation

Usually, the goal when determining features is to discover the *factors of variation* that explain the observed data. For example, factors of variation for a speech recording include speaker age, sex, and accent. It can be difficult to extract high level features, though. Recognizing speaker accent itself requires nearly human-level understanding.

Deep learning solves this problem by expressing representations in terms of other, simpler ones. One such model is the *multilayer perceptron*, which is a function (composed of simpler functions) mapping input values to output values. This gives a first perspective on deep learning:

```
A series of mathematical functions each provide a new
representation of the data
```

Alternatively, it can be thought of more like a state machine:

```
Each layer of depth allows the computer to learn a multistep
computer program. Each layer captures the state after executing
a set of instructions
```

According to the second view, not all information in each layer necessarily encodes information about factors of variation – some will encode state information.

The depth of a model is measured in two ways

1. In terms of the number of sequential instructions that must be executed to evaluate the architecture ("longest path").

   This path length will vary depending on the functions we allow individual nodes to apply.

2. In terms of the depth of the graph describing concept relations

   Because understanding of