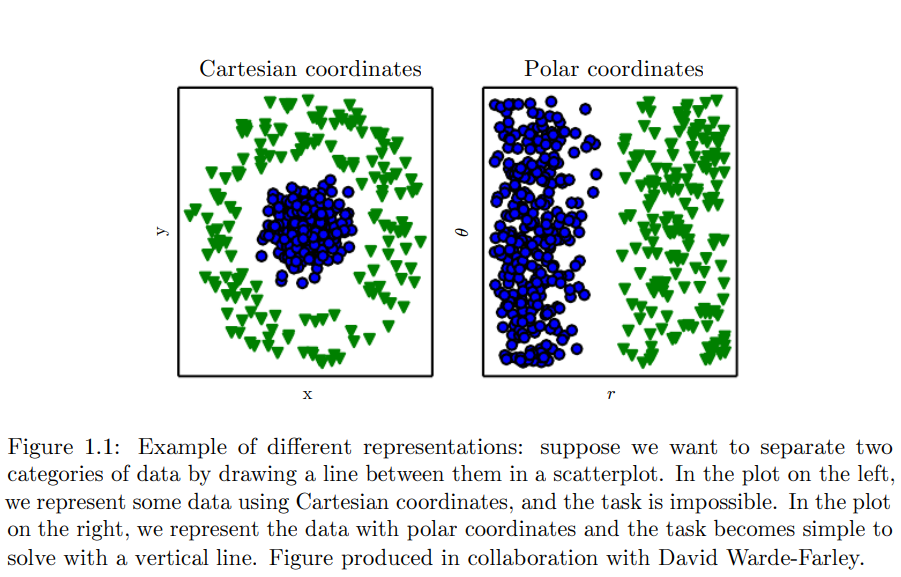
Ian Goodfellow – Deep Learning

# Introduction

Machine learning algorithms **logistic regression** can recommend cesarean delivery(kejsarsnitt), **naive Bayes** can separate spam e-mails.

The performance depends heavily on the **representation** of the data that is given. The information is, in simple ML cases, presented as **features**.

To choose the data representation, indexation and features carefully is really important. Some tasks may be unsolvable or take exponentially longer time if chosen wrongly. See picture below. 

When using **representation learning** you can detect the representation itself, hence not needing any predefined features. Usually renders better performance than hand designed representations. The quintessential example of a representation learning algorithm is the **autoencoder**. They are trained to keep as much info as possible but also make the new representation have various nice properties.

When designing features and algorithms we want to separate the **factors of variation**.

The quintessential example of a deep learning model is the feedforward deep network or **multilayer perceptron (MLP)**. A multilayer perceptron is just a mathematical function mapping some set of input values to output values.

Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones

# Historical trends in deep learning

## The many names and changing fortunes of neural networks

Has existed since 1940s.

Three waved of development. **Cybernetics** in 40s-60s. **Connectionism** in 80s-90s. **Deep learning or artificial neural network (ANN)** starting 2006.

**Adaptive linear element (ADALINE)** could predict a real number which it returned. The weights were adjusted with a special case of an algorithm called **stochastic gradient descent**. Models based on the f(**x, w**) used by ADALINE are called **linear models**.

Most neural networks today are based on a model neuron called the **rectified linear unit**.

Several key concepts arose during the connectionism or **parallel distributed processing** in the 1980s. One of these is that of **distributed representation.** The idea is that each input to a system should be represented by many features, and each feature should be involved in the representation of many possible inputs. For example, a vision system. Instead of having a neron for red car, red bird, red bike, green car, green bird… we would have one layer describing the object identity and one describing the color.

Another key concept found was the one of back-propagation to train deep neural networks. This is currently the most popular approach to training deep models.

## Increasing dataset sizes

A rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around 5 000 labeled examples per category, and will match or exceed human performance when trained with a dataset containing at least 10 million labeled examples. Working with datasets smaller than this is an important research area.

## Increasing model sizes

Amount of neurons in the ANN is doubled every 2.4 years. If this continues we will have the same as a human brain in 2050s.

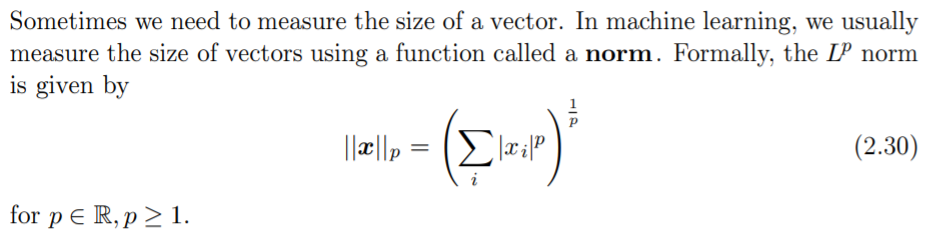
### Increasing accuracy, complexity and real-world impact

CNN surpassed opponents in 2012 in image recognition, now it is dominating the discipline. It has yielded superhuman performance in traffic sign detection.

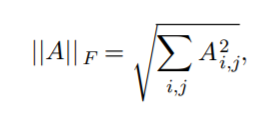
**Reinforcement learning** is a form of unsupervised learning which is useful in learning to play e.g. Atari games or control robots.

Applied math and machine learning basics

# Norms



**Euclidean norm** is the Euclidean distance from the origin to the point identified by x, also called the L2 norm.

**Frobenius norm** measures the size of a matrix. 

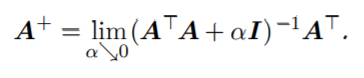
# Singular value decomposition

**Singular value decomposition (SVD)** is kind of the same as an eigenvalue except it works for non-quadratic matrices.

# The Moore-Penrose pseudoinverse

When the matrix is not quadradic but needs to be solved anyway.

The Moore-Penrose pseudoinverse allows us to make some headway in these cases. The pseudoinverse of A is defined as a matrix



Practical algorithms for computing the pseudoinverse are not based on this definition, but rather the formula

