D2L

1. Introduction

When you are able to devise solutions that work 100% of the time using first principle programming, you should not be using machine learning.

Sometimes the program that we are looking for follows a pattern that changes over time, and we need our programs to adapt. In other cases, the relationship (say between pixels, and abstract categories) may be too complicated, requiring thousands or millions of computations that are beyond our conscious understanding (even if our eyes manage the task effortlessly).

ML is the study of powerful techniques that can learn from experience.

As an ML algorithm accumulates more experience, typically in the form of observational data or interactions with an environment, its performance improves. Contrast this with our deterministic e-commerce platform, which performs according to the same business logic, no matter how much experience accrues, until the developers themselves learn and decide that it is time to update the software.

* 1. Motivating Example

In the ML approach, we do not attempt to design a system explicitly do the task. Instead, we define a flexible program whose behavior is determined by a number of parameters. Then we use the dataset to determine the best possible set of parameters, those that improve the performance of our program with respect to some measure of performance on the task of interest.

In deep learning, the learning is the process by which we discover the right setting of the knobs (model parameters) coercing the desired behavior from our model.

Deep models are deep in precisely the sense that they learn many layers of computation. It turns out that these many-layered (or hierarchical) models are capable of addressing low-level perceptual data in a way that previous tools could not. One key advantage of deep learning is that it replaces not only the shallow models at the end of traditional learning pipelines, but also the labor-intensive process of feature engineering.

* 1. Key components
     1. Data

When every example is characterized by the same number of numerical values, we say that the data consists of fixed-length vectors and we describe the (constant) length of the vectors as the dimensionality of the data.

One major advantage of deep learning over traditional methods is the comparative grace with which modern models can handle varying-length data.

Many of the most exciting models in deep learning do not work without large datasets. Some others work in the low-data regime, but are no better than traditional approaches.

* + 1. Models

Deep learning is differentiated from classical approaches principally by the set of powerful models that it focuses on. These models consist of many successive transformations of the data that are chained together top to bottom, thus the name deep learning.

* + 1. Objective Functions

In order to develop a formal mathematical system of learning machines, we need to have formal measures of how good (or bad) our models are. In machine learning, and optimization more generally, we call these objective functions.

Typically, the loss function is defined with respect to the model’s parameters and depends upon the dataset. The best values of our model’s parameters are learned by minimizing the loss incurred on a training set consisting of some number of examples collected for training. However, doing well on the training data does not guarantee that we will do well on (unseen) test data.

* 1. Kinds of ML
     1. Supervised Learning

The supervision comes into play because for choosing the parameters θθ, we (the supervisors) provide the model with a dataset consisting of labeled examples (xi,yixi,yi), where each example xixi is matched with the correct label.

In probabilistic terms, we typically are interested in estimating the conditional probability P(y|x)P(y|x).

