Machine Learning: Foundations

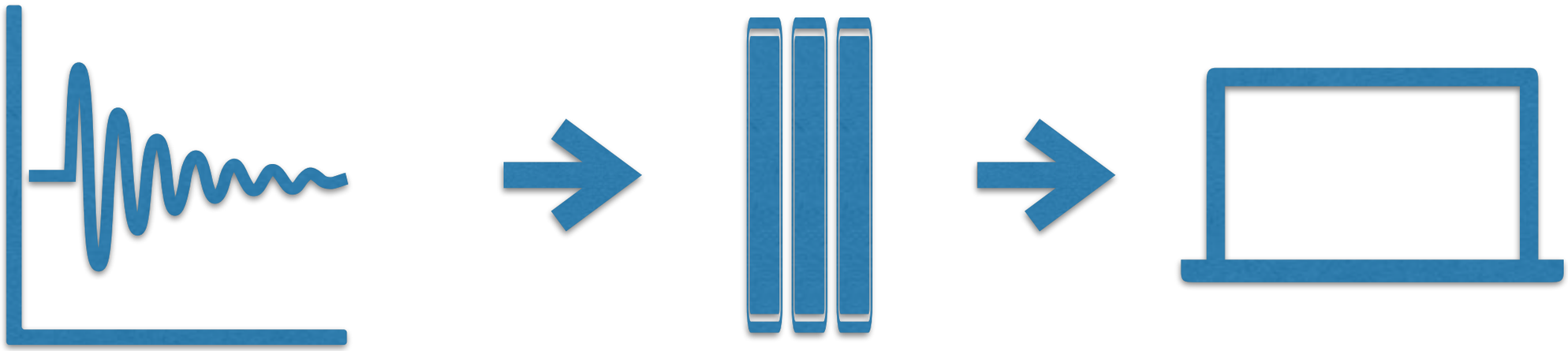
Introduction to Neural Networks

SUVRIT SRA

Massachusetts Institute of Technology

ml.mit.edu

Introduction

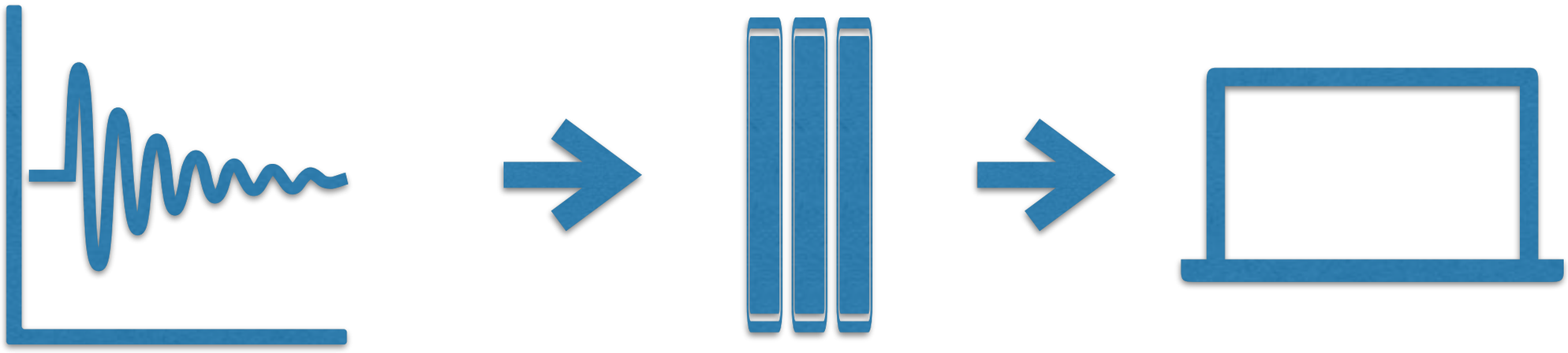


See for example

https://scikit-learn.org/stable/modules/feature_extraction.html

Exercise: Try feature extraction to turn an email into a vector

Introduction



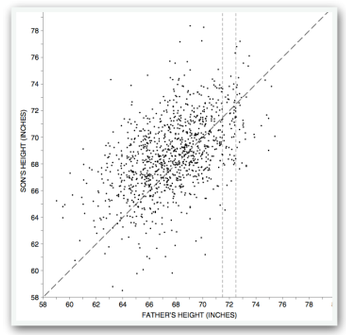
The machine learning mindset

What if we learn the features from data instead of arduous feature extraction?

Simple view of neural network models

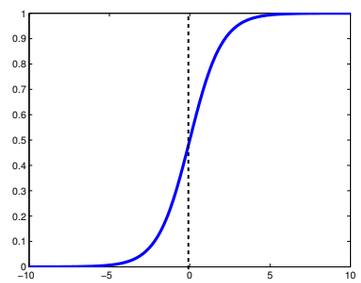
linear model

$$h(x) = w^T x + b$$



linear model+
sigmoid activation
(logistic regression)

$$h(x) = \sigma(w^T x + b)$$



How to get a classifier out of logistic regression h(x) above?

$$y_{\text{predicted}} = \begin{cases} 0 & \text{if } h(x) < \frac{1}{2}, \\ 1 & \text{if } h(x) \geq \frac{1}{2} \end{cases}$$

for multiclass classification, use the 'softmax' rule to do the classification

Simple view of neural network models

Key idea

- 'm' different classifiers / experts, leading to a vector of predictions

$$x \mapsto \big(\sigma(w_1^T x + b_1), \ldots, \sigma(w_m^T x + b_m)\big)$$
$$\equiv \sigma(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b})$$

use logistic model again to learn how to combine these into one prediction!

$$\boldsymbol{x} \mapsto \sigma(\boldsymbol{u}^T \sigma(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}_1) + c)$$

- continue this process recursively to obtain a "deep model"

$$\mathcal{H} = x \mapsto \sigma_L(W_L \sigma_{L-1}(W_{L-1} \cdots \sigma_1(W_1 x + b_1) \cdots + b_{L-1}) + b_L)$$

as our mathematical model. This is called a ***feed-forward deep neural network***

wow, that's it?!

Let us look more slowly….