Topic 1 - Introduction

Machine Learning

- Supervised Learning

Given: (x_i, y_i), $i = 1, 2, \dots, m$

x_i: Input; y_i: Output/Label

With this training set, our goal is then to find the relationship between the input and the output.

Find a function f s.t. : $f(x_i) \simeq y_i$, $i = 1, \dots, m$.

If a new input x comes in, we use $f(x)$ as the prediction of the prediction of the label of x.

This is a typical supervised (learning) task.

This problem is called a data-fitting problem.

Example: Handwritten digits

- Unsupervised Learning

Only input data $\{x_i\}_{i=1}^m$ without labels!

Here we may be required to find the common features or some essential parameters in those x_i. If this happens, this problem is known as the dimension reduction.

Example: Given some pictures of human faces $\{x_i\}_{i=1}^m$
Generate some "artificial" faces?
We can use GAN or VAE.

What is Learning?

Learning = Representation + Evaluation + Optimization

Representation:

① How to represent the input data x_i?

- E.g. Matrices represents images (entries ⟷ pixel)
- Vectors rep. stock prices (entries ⟷ stock price at a time)

② How to rep. the function f?
Which function f should we use?

The set of all "good" functions is called "Hypothesis Space".

- E.g. we may choose f in

  - the set of "linear functions" → are completely predictable!
  - the set of functions defined by "deep neural networks"

- Math tools: "the space of functions"
  ↳ "functional analysis"

Evaluation:

① → Which function best fits our machine learning task. Or how to define "the best" function in the hypothesis space?

Need to define a function that maps a function in the hypothesis space to a number.

"function of function"

= functional ⇝ maths tool: Functional analysis

Example: $F: C([0,1]) \to \mathbb{R}$
$F(f) = f(0.5) \quad \forall f \in C([0,1])$

The Evaluation process needs domain knowledge.

② How to define "the best" representation of the input data?

For example: In image data, one representation is the so-called sparse representation.

Math tools: Harmonic analysis / Fourier analysis

Optimization:

We need to minimize the functional. We need a numerical solver to get the optimal solution numerically.

- Convex optimizations, local minima is global minima!
- Nowadays non-convex optimization becomes more and more important, for example in Deep Learning.