

Math 3332 Data Analytic Tools

- Introduce **Mathematical analysis tools** for data analysis.

- Data analysis / Machine Learning

Supervised Learning:

Given input-output pairs (x_i, y_i) $i=1, 2, \dots, m$.

output (e.g., a label of
dog or not dog)

↓

↑

input (e.g. an image)

find an f s.t. $f(x_i) \approx y_i$. $\forall i=1, 2, \dots, m$.

Therefore, given a new input x , we can predict the output y by $y \approx f(x)$.

Unsupervised Learning:

Only data input.

For example, given many photos of human faces,
can we produce a new virtual face that looks natural but doesn't belong to a real person?

Many other tasks.

- **Learning = Representation + Evaluation + Optimization** (Pedro Domingos)

Representation: ① How to represent a learner? Which set should a learner be in?

This set is called the **hypothesis space** of the learner.

— related tools are "space of functions"

② How to represent the input?

— candidate tools: vectors, graphs, manifolds, ...

Evaluation: ① How to pick the best learner from the hypothesis space?

— Needs "calculus of (functions of functions)"

② How to represent the input effectively? → Called functionals

— Needs Linear algebra, graph theory, manifold calculus, etc
harmonic analysis

Optimization: Numerical solvers.

Many of the resulting optimization is **convex optimization**.
Therefore, we need "convex analysis" and "convex optimization".

- This course: ① Basic functional analysis (calculus of functionals)
- ② Basic convex analysis
- ③ Fourier analysis, Wavelet analysis if time allowed
(for efficient representation of image/video/signal input)