Lecture 1: Basics of ML.

⇒ Theme: theoretical foundations of ML.

↓ feature learning ↓ AI safety.

⇒ Importance: We want to understand how & why ML works.

- Train a model ⇒ test on a new task
  - → knowledge doesn't transfer
  - → not enough data

  } fundamental statistical problems.

  - → computational issue → optimization issue.

  (objective function is flawed)

- guide technical decision
- reduce trial and error
- forecast outcomes and risks.
- inspire new methods.

} esp. useful demand / supplies resource is large

⇒ Structure and logic of this course.

ML (AI in general) many success stories.

$\boxed{\text{Chatbot}}$. $\boxed{\text{Medical diagnose}}$ $\boxed{\text{Alpha-Go}}$.

text image RL. environments' specific structures.

We don't care about data modalities, have a general view high-dimensional samples, generated from a certain distribution.

$$\mathbb{X} = \{x_1, \cdots x_n\}. \qquad x_i \sim P_X.$$

$$\mathbb{Z} = \{(x_1,y_1), \cdots (x_n,y_n)\} \qquad x_i, y_i \sim P_{X,Y}.$$

Basics of ML. consists of the following elements.

New era (AGI)

- data (MNIST. CIFAR, IMAGENET)

wiki, crawled from the Internet.

- loss function (measuring the difference of model's prediction vs y_i: l_2, cross-entropy, etc.

next word prediction.

$-P(x_{t+1} | x_1 - x_t)$.

- model (linear/affine, kernel, neural networks.

transformer, Bert.

- optimization alg. (SGD, Adam, adagrad). etc

—

- risks on AI safety

Security: adversarial attack

poisonous attack

privacy: reidentification.

clean data poisonous attack.

copyright

Chatbot (finetuned w/

your dialog.

alignment.

⇒. Supervised learning.

Input data $\{(x_i, y_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$.

input data, feature

label response.

or $\mathbb{R}^c$.

Goal: Find prediction function $f_\theta : \mathbb{R}^d \to \mathbb{R}$ $(\mathbb{R}^d \to \mathbb{R}^c)$.

for multi class.

such that $f_\theta(x_i) \simeq y_i, \ \forall i$.

parametric model.

Empirical Risk Minimization:

$$\hat{R}_n := \min_\theta \frac{1}{n} \sum_{i=1}^n \ell(f_\theta(x_i), y_i)$$

$\Downarrow$ concentrate

population.

$$R = \mathbb{E}_{(x,y)\sim P_{x,y}} \ell(f_\theta(x), y).$$

Hope:

f_θ s.t

$\hat{R}_n(f_\theta)$ small

$\Downarrow$

$R(f_\theta)$ small.

⇒ More types of ML tasks:

Traditional types of ML:

- supervised learning
- unsupervised learning.
- Reinforcement learning.

Spectrum between supervised ⇒ unsup. $\mathcal{L}$:

More labeled data involved

- unsupervised $\mathcal{L}$.
- self-supervised.
- meta-learning.
- domain generalization.
- semi-supervised learning
- supervised learning.

feature learning.

$\mathcal{L}$: label data. U: unlabeled data.

S: source data.

T: target data.

e: # enviroments or tasks.

Learning tasks	data that model trained on	data model is tested on
self-supervised. (un)	U^S, L^T few-shot	U^T
meta-learning.	$L_1^S, L_2^S, \dots L_e^S$, L^T	U^T
domain generalization.	$L_1^S, L_2^S, \dots L_e^S$	U^T.
semi-supervised learning	L, U	U
supervised learning.	L	U.
reinforcement learning	Source environment	target environment.

⇒ Theoretical groundings of ML algorithms in general:

$\circledast$ objective function

supervised	unsupervised	privacy attack.
$\sum_{i=1} \ell(f_\theta(x_i), y_i)$	$\min_X \lVert M - X X^T \rVert_F^2$	$\min_X \ell(O, F(X)) + \text{prior}(X)$. $(*)$

LT requires studying the two aspects;

$$\sum_{(i,j)\in\Omega} \lVert M_{ij} - (XY^T)_{ij} \rVert_F^2$$

statistics ⇓ whether the optimal solution generalizes.

optimization: ↓ whether our algorithm can find a global minimum.

when is the optimal of $(*)$ close to the ground-truth.

together they form Learning theory.