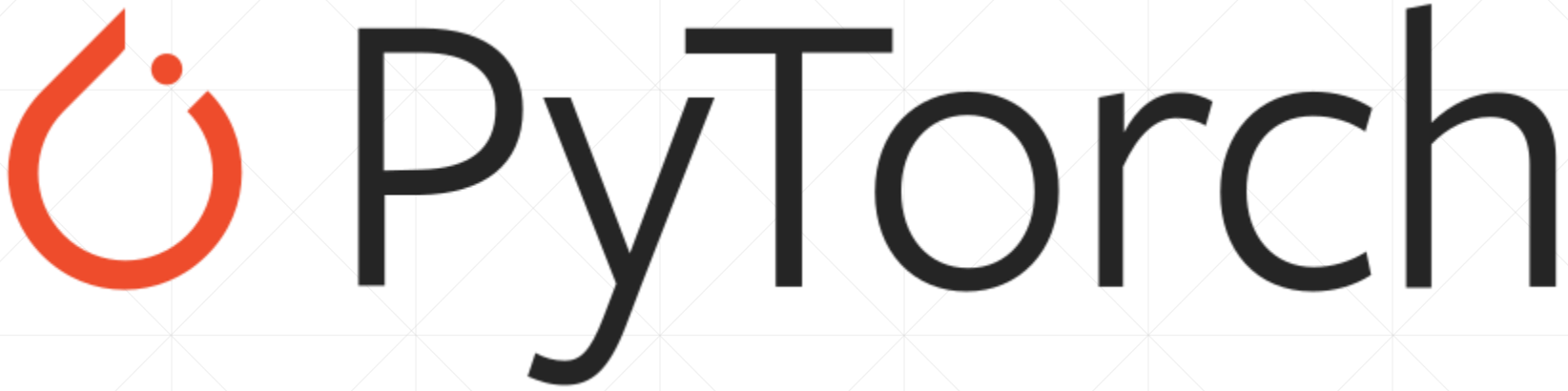


你好，MNIST

主讲人：龙良曲

DL is NOT a Toy

- MNIST
  - each number owns 7000 images
  - train/test splitting: 60k vs 10k

NO deep learning, just function mapping

- $X = [v1, v2, ..., v784]$
  - X: [1, dx]
- $H1 = XW1 + b1$
  - W1: [d1, dx]
  - b1: [d1]
- $H2 = H1W2 + b2$
  - W2: [d2, d1]
  - b2: [d2]
- $H3 = H2W3 + b3$
  - W3: [10, d2]
  - b3: [10]

Loss?

- H3: [1, d3]
- Y: [0/1/.../9]
  - eg.: 1 => [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
  - eg.: 3 => [0, 0, 0, 3, 0, 0, 0, 0, 0, 0]
- Euclidean Distance: H3 vs Y

In a nutshell

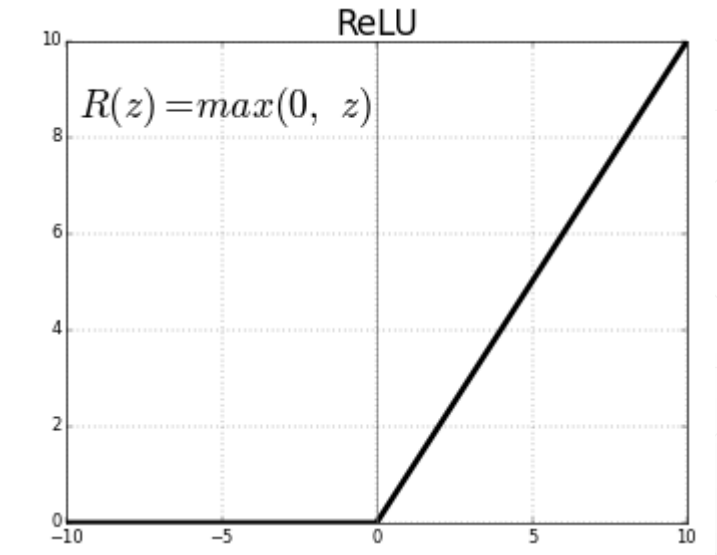
- $pred = W_3 * \{W_2[W_1X + b_1] + b_2\} + b_3$

- Linear Combination?

Non-linear Factor

- ReLU

- H1 = **relu**(XW1 + b1)
- H2 =**relu**(H1W2 + b2)
- H3 =**relu**(H2W3 + 3)



- $pred = W_3 * \{W_2[W_1X + b_1] + b_2\} + b_3$

Gradient Descent

- objective $= \sum(pred - Y)^2$
- minimize objective
- [W1, W2, W3]
- [b1, b2, b3]

Inference

- For a new X
- $pred = W_3 * \{W_2[W_1X + b_1] + b_2\} + b_3$
  - [0.1, 0.8, 0.01, 0, 0.02, …]
- $argmax(pred)$

下一课时

实战MNIST

Thank You.