

STEPS TO DESIGN A MODEL

This assume you already has a clean set of data

Step 1: Looking at your data

- Look at a random subset of your data
- Look at the extreme:
 - Largest/smallest file size

Step 2: Try solve the task manually

- Can you solve the task yourself?
- What is the intuition you used to solve it? Focus on what is hard for you when solving the task?

Step 3: Feed your data to the network

- Start with the simplest possible architecture, make sure the input/output is correct.
- Normalize input before feeding into the network.

Why normalize?

- To remove the dependency of weights to each other with respect to the loss.

- Consider these 2 scenarios:

Given input $x = [2, 3]$, $w = [w_1, w_2]$ and output y (scalar):

- Not normalized:

$$\nabla_{w_1} L_2 = \epsilon \times 2, \text{ where } \epsilon = \nabla_{\hat{y}} L_2$$

$$\nabla_{w_2} L_2 = \epsilon \times 3 = -(y - \hat{y})$$

$\Rightarrow w_2$ is forced to change 1.5 times more than w_1 , they are locked in this 2:3 ratio

- Normalized: $x_{\text{norm}} = [-0.5, 0.5]$

$$\nabla_{w_1} L_2 = \epsilon \times (-0.5)$$

$$\nabla_{w_2} L_2 = \epsilon \times 0.5$$

$\Rightarrow w_2$ and w_1 now change more independently, no more fixed ratio.

Step 4: Design your network

- Start simple