

①

Few shot meta-learning  
few shot meta learning?

Simple view: Optimal model params

$$\theta^* = \arg \min_{\theta} \underbrace{E_{D \sim P(D)}}_{\text{Dataset}} [L_{\theta}(D)]$$

one dataset as one data sample??

$$D = \langle S, B \rangle \quad \begin{array}{l} \nearrow \text{Prediction} \\ \downarrow \text{Learning} \end{array} \quad [K \text{ shot - } N\text{-class classification}]$$

Training in the same way as Testing!

$$D = \{(x_i, y_i)\} \in L^{\text{Label}}$$

classifier  $f_{\theta}$

output probabilities for  $y | x \rightarrow P_{\theta}(y | x)$

Optimal Parameters

$$\theta^* = \arg \max_{\theta} E_{(x,y) \sim D} [P_{\theta}(y | x)]$$

$$\theta^* = \arg \max_{\theta} E_{B \sim D} \left[ \sum_{x,y \in B} P_{\theta}(y | x) \right]$$

// Expectation distribution changes.

Few Dataset  $\rightarrow$  small support set  $\rightarrow$  fake  $\rightarrow$  fast learning.

(ii)

steps

(i) subset of labels  $L \subset \mathcal{L}^{\text{label}}$ . taking few labels. [2 out of 6 classes] maybe

(ii) sample support  $s^L \subset \mathcal{D}$ , training batch  $B^L \subset \mathcal{D}$

$$y \in L, \forall (x, y) \in s^L, B^L$$

(iii) support set  $\rightarrow$  part of model input.

(iv) optimization uses mini batch  $B_L$

$(\underline{s}^L, \underline{B}^L) \rightarrow$  one data point??

model trained to generalized to other dataset.

$$\theta = \underset{\theta}{\operatorname{argmax}} \mathbb{E}_{L \subset \mathcal{L}} \left[ \mathbb{E}_{s^L \subset \mathcal{D}, B^L \subset \mathcal{D}} \left[ \sum_{(x, y) \in B^L} P_{\theta}(x, y, s^L) \right] \right]$$

[meta learning Params.]

[optimization: to be good at many]

Learner Vs meta-learner:

Two stage updates:

(i)  $f_{\theta} \rightarrow$  learners model.

(ii) update via support set  $s$ ;  $\theta' = g_{\phi}(\theta, s)$

$$\mathbb{E}_{L \subset \mathcal{L}} \left[ \mathbb{E}_{s^L \subset \mathcal{D}, B^L \subset \mathcal{D}} \left[ \sum_{(x, y) \in B^L} P_{g_{\phi}(\theta, s^L)}(y | x) \right] \right]$$