Learning from very few data

Few sample Learning (FSL)

a single or a handful of examples !

pre frontal cortex (PFC) : prominent human working memory in human brain: learning ability

Mapping function :

$$f \in \mathcal{F} : x \rightarrow y$$

$$D_t = \{(x_i, y_i)\}_{i=1}^{n}$$

joint distribution $P_{x \times y}$

$$\mathcal{E}_{ex} = \mathbb{E}_{(x,y) \sim P_{x \times y}} L(f(x), y)$$

$\downarrow$ $P_{x \times y}$ is unknown

$$\mathcal{E}_{ex} = \mathbb{E}_{(x,y)\sim D_t} L(f(x), y)$$

Generalization Error:

$$\mathcal{E} = |\mathcal{E}_{ex} - \mathcal{E}_{em}|$$

$\mathcal{E}$ is big since the Function Space $\mathcal{F}$ is large

$$\min_f \mathcal{E}_{em}, \quad \text{s.t.} \begin{cases} f(x_i) = y_i \\ \forall (x_i, y_i) \in D_t \end{cases}$$

regularization on the function f

Compress the redundant optional space of function f and thereby reduce its generalization

error.

E. G. Miller (2000) Congealing Algorithm

G. Koch (2015) DL + FSL

Before 2015 (Non-DL):

Generative Model

estimate { joint distribution $P(x, y)$
conditional distribution $P(x|y)$

Congealing Algorithm

Variational Bayesian Framework (VBF)

Bayesian Program Learning (BPF)

After 2015 (DL):

- Data Augmentation
- Metric Learning
- Meta Learning

Siamese CNN

Matching Nets

MAML

Meta-Learner LSTM

MANN

MetaNet

Prototypical Nets

LGM-Nets

$$D_T = \{ D_{trn}, D_{tst} \}$$

$$\begin{cases} D_{trn} = \{ (x_i, y_i) \}_{i=1}^{N_{trn}} \\ D_{tst} = \{ x_j \}_{j=1}^{N_{tst}} \end{cases}$$

$$\begin{cases} x_i, x_j \in X_T \subset X \\ y_i \in Y_T \subset Y \end{cases}$$

$$D_T = \{ X_T, P(X_T) \}$$

$P(X_T)$: Marginal probability distribution

C task classes

k (very small, 1, 5) samples per class

$N_{trn} = Ck$ (C-way k-shot task)

Goal: produce a target predictive function

$$f \in \mathcal{F} : X \rightarrow Y$$

which can make predictions for ~~test~~ samples in D_{tst}.

$$D_A = \{(x_i^a, y_i^a)\}_{i=1}^{N_{aux}}$$

$$x_i^a \in X_A \in X, \quad y_i^a \in Y_A \subset Y$$