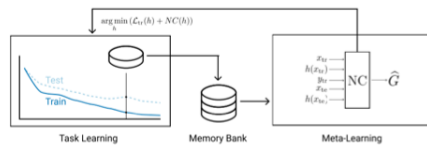


Meta-learning to Predict Generalization

Reference - *Neural Complexity Measures*

During meta-learning tasks include train and test data: the gap between train/test loss is available. NC trains another network to predict this generalization gap



$$\mathcal{L}_{NC} = \mathcal{L}_{Test} - \mathcal{L}_{Train} - NC(h)$$

$$Q = f(X_{te}), K = f(X_{tr}), V = [K; Y_{tr}]$$

$$NC(X_{tr}, X_{te}, Y_{tr}, h(X_{tr}), h(X_{te})) = \frac{1}{m'} \sum_{i=1}^{m'} g(A_i)$$

$$\text{where } A = \frac{\text{Softmax}(QK^T)}{\sqrt{d}}V$$

NC-regularized loss: $\mathcal{L}_{reg} = \mathcal{L}_{Test} + \lambda NC(h)$; λ increased gradually over tasks.

↑
L_{NC}

$f(x_{te})$

$[f(x_{tr}); y_{tr}]$



$g(\cdot) \rightarrow \text{scalar}$