Stochastic Multiple Choice Learning For Training Diverse Deep Ensembles

Stefan Lee1, Senthil Purushwalkam2, Michael Cogswell3, Viresh Ranjan4, David Crandall5, Dhruv Batra2
Virginia Tech1 Georgia Tech2 Carnegie Mellon University3 Indiana University4

1 OVERVIEW: The Need for Diversity
Many interesting inference problems have some degree of ambiguity, often as an implicit property of an uncertain world.

In the face of ambiguity, humans can give multiple likely answers to articulate multimodal beliefs. One natural method to generate multiple outputs is to train an ensemble of models; however, we find independently trained networks typically produce similar outputs.

2 STOCHASTIC MULTIPLE CHOICE LEARNING (SMCL)
To encourage the specialization of ensemble members, we consider a loss with respect to a single model’s output

\[ L_\text{O} (D) = \sum_{m=1}^{M} \min_{e \in \mathbb{M}} \ell (y, f_m (x)) = \sum_{m=1}^{M} p_m \ell (y, f_m (x)) \]

where \( p_m \) is if model \( m \) has the lowest loss on example \( e \) and 0 otherwise. Holding \( p_m \) fixed, the gradient with respect to a single model’s output \( f_m (x) \)

\[ \frac{\partial L_\text{O}}{\partial f_m} = p_m \frac{\partial \ell (y, f_m (x))}{\partial f_m} \]

Leads to a simple training algorithm to minimize the oracle loss in SGD-based learners which we call Stochastic Multiple Choice Learning (SMCL).

SMCL Training Algorithm:
For each example in a batch:
1) Compute the loss of the example for each model in the ensemble.
2) Back-propagate the gradient only to the model with lowest loss.

This ‘Winner-Take-Gradient’ training is agnostic to both model architecture and loss.

3 SPECIALIZATION IN IMAGE CLASSIFICATION
To test SMCL in a simple setting, we train ensembles on CIFAR10 using a small CNN model. We find sharp, class-based specializations emerge in SMCL trained ensembles.

4 SPECIALIZATION IN SEMANTIC SEGMENTATION

5 SPECIALIZATION IN IMAGE CAPTIONING
We evaluate on the MSCOCO image caption task, training ensembles of the CNN+LSTM model of Karpathy et al. (2015, with and without CNN fine-tuning.

6 CONCLUSION
For many complex inference tasks, there is implicit ambiguity and/or multiple correct possible outputs. By directly optimizing for the oracle loss, our SMCL allows an ensemble to specialize in response to ambiguity and multimodal outputs distributions.

Stochastic Multiple Choice Learning is easy to implement effective, and model agnostic.