Problem

Transferability estimation

Estimating how easy it is to transfer knowledge from one classification task to another

- Given a pre-trained source model and a target data set
- Develop a measure (a score) for how effectively transfer learning can transfer from the source model to the target data
- Transferability measure should be easy and cheap to compute $\rightarrow$ ideally without training

Why do we need transferability estimation?

- Help understand the relationships/structures between tasks
- Select groups of highly transferable tasks for joint training
- Select good source models for transfer learning
  - Potentially reduce training data size and training time

Our contributions

- We develop a novel transferability measure, Log Expected Empirical Prediction (LEEP), for deep networks
- Properties of LEEP:
  - Very simple
  - **Clear interpretation**: average log-likelihood of the expected empirical predictor
  - **Easy to compute**: no training needed, only requires one forward pass through target data set
  - Can be applied to most modern deep networks

Log Expected Empirical Prediction (LEEP) (1)

- Assume source model θ and target data set $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
- We compute LEEP score between θ and $\mathcal{D}$ in 3 steps.

1. Apply θ to each input x_i to get dummy label distribution $\theta(x_i)$.
   - $\theta(x_i)$ is a distribution on source label set $\mathcal{Z}$
   - Labels in $\mathcal{Z}$ may not semantically relate to true label y_i of x_i e.g., $\mathcal{Z}$ is ImageNet labels but (x_i, y_i) is from CIFAR

2. Compute empirical conditional distribution of target label y given dummy source label z

   Empirical joint dist: $\hat{P}(y, z) = \sum_{i: y_i = y} \theta(x_i)_z / n$

   Empirical marginal dist: $\hat{P}(z) = \sum_y \hat{P}(y, z)$

   Empirical conditional dist: $\hat{P}(y|z) = \hat{P}(y, z) / \hat{P}(z)$

Log Expected Empirical Prediction (LEEP) (2)

Expected Empirical Predictor (EEP)

A classifier that predicts the label y of an input x as follows:

- First, randomly drawing a dummy label z from $\theta(x)$
- Then, randomly drawing y from $\hat{P}(y|z)$

Equivalently, $y \sim \sum_z \hat{P}(y|z)\ \theta(x)_z$

3. LEEP is the average log-likelihood of EEP given data $\mathcal{D}$:

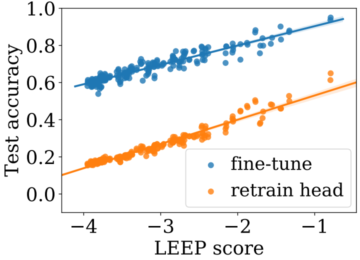
$$T(\theta, \mathcal{D}) = \frac{1}{n} \sum_i \log \left(\sum_z \hat{P}(y_i|z)\ \theta(x_i)_z \right)$$

Experiment: overview

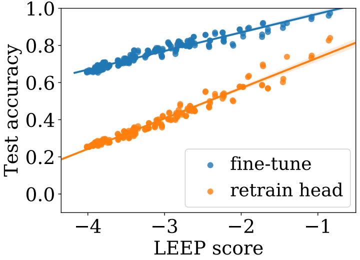
- Aim: show that LEEP can predict actual transfer accuracy
- Procedure:
  - Consider many random transfer learning tasks
  - Compute LEEP scores for these tasks
  - Compute actual test accuracy of transfer learning methods on these tasks
  - Evaluate correlations between LEEP scores and the test accuracies
- Transfer methods:
  - **Retrain head**: only retrain last fully connected layer using target set
  - **Fine-tune**: replace the head classifier and fine-tune all model parameters with SGD

Experiment: LEEP vs. Transfer Accuracy

- Compare LEEP score with test accuracy of transferred models on 200 random target tasks
- Result: LEEP scores highly correlated with actual test accuracies (correlation coefficients > 0.94)



ImageNet $\rightarrow$ CIFAR100 (ResNet18)



CIFAR10 $\rightarrow$ CIFAR100 (ResNet20)