Chapter19 Transfer Learning

1. Data not directly related to the task considered
   a) Similar domain, different tasks
   b) Similar task, different domains
   c) A large amount of source data and very little target data
2. Model Fine-tuning
   a) Source data and target data are labeled
   b) Train a model by source data, and then fine-tune the model by target data
      Only limited target data, so be careful about overfitting
   c) Example: Speaker Adaption
   d) One-shot learning: only a few examples in target domain
   e) Conservative learning
      Initialize the model by source data and then train it by target data
   f) Layer Transfer
      Copy some layers, only train the rest layers
      In speech domain, usually copy the last few layers
      In image domain, usually copy the first few layers
3. Multitask Learning
   a) Source data and target data are labeled
   b) Example: Multilingual Speech Recognition
   c) Progressive Neural Network
4. Domain Adversarial Training
   a) Source data is labeled, whereas target data is unlabeled
   b) Similar to GAN
   c) Feature extractor: maximize label classification accuracy, minimize domain classification accuracy
   d) Label predictor: maximize label classification accuracy
   e) Domain classifier: maximize domain classification accuracy
5. Zero-shot Learning
   a) Source data is labeled, whereas target data is unlabeled
   b) In speech recognition, we can't have all possible words in the source data
   c) Representing each class by its attribute
   d) Attribute Embedding: f(x) and g(y) as close as possible

$$f^*, g^* = argmin \sum_n \max\left(0, k - f(x^n)g(y^n) + \max\left(f(x^n)g(y^m)\right)\right)$$

k is a margin defined by programmer

Zero loss: $k - f(x^n)g(y^n) + \max\left(f(x^n)g(y^m)\right) < 0$

$$f(x^n)g(y^n) - \max\left(f(x^n)g(y^m)\right) > k$$

6. Self-taught Learning: Source data is unlabeled, whereas target data is labeled
7. Self-taught Clustering: Source data and target data are unlabeled